Writing as a conceptual process

Pool, Elisabeth Mathilde Catharina van der

Publication date:
1995

Document Version
Publisher's PDF, also known as Version of record

WRITING

as a conceptual process

A text-analytical study of developmental aspects

Els van der Pool

Writing as a conceptual process

A text-analytical study of developmental aspects

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Brabant,
op gezag van de rector magnificus,
prof. dr L.F.W. de Klerk,
in het openbaar te verdedigen
ten overstaan van een
door het college van dekanen
aangewezen commissie
in de aula van de Universiteit
op vrijdag 22 september 1995
om 14.15 uur precies

door

Elisabeth Mathilde Catharina van der Pool

geboren op 11 maart 1965 te Breda

Promotor: Prof. dr L.G.M. Noordman

Co-promotor: Dr C.H. van Wijk

Writing as a conceptual process

A text-analytical study of developmental aspects

(met een samenvatting in het Nederlands)

Els van der Pool

Pool, E.M.C. van der

Writing as a conceptual process. A text-analytical study of developmental aspects
Elisabeth Mathilde Catharina van der Pool
Proefschrift Katholieke Unviversiteit Brabant, Tilburg.
Met lit.opg. - Met een samenvatting in het Nederlands.

ISBN 90-9008573-4
NUGI 941

Trefw.: Tekstwetenschap, schrijfonderzoek, taalontwikkeling

Nederlandse titel: Schrijven als conceptueel proces.
Een tekst-analytische studie van ontwikkelingsaspecten.

Cover designed by Marleen van Heck and Els van der Pool

Printed on 10 pt Times Roman; printed by Copyprint 2000, Enschede

to my mother

Preface

A research project is like finding your way through a labyrinth: the path may have many branches, it may be a long and winding road before you reach your goal, and there may even be dead ends. A project like this can only become a success with the support of many people.

First of all, I would like to thank Carel van Wijk and Leo Noordman. In his great enthusiasm for natural language generation, Carel taught me how to do research in this field. We had many lively discussions about designing studies, modelling the writing process, and about the appendix of chapter 4; they have all been crucial for this thesis. We also had very funny discussions about, among other things, the use of WP4.2 and collecting flippos; they were crucial for our pleasant cooperation. Leo also guided my project critically and constructively. He stimulated me in delineating and specifying my ideas and statements. I could always rely on both Carel's and Leo's expertise and experience.

I enjoyed the fruitful discussions - live and by e-mail - with Joost Schilperoord. They were very inspiring and of great value for this study. It was a pleasure to work with a peer in such a humorous way. I also enjoyed the discussions I had with my roommate José Sanders. Both of us being PhD students, we faced similar problems and she could help me find solutions.

For the interest in and discussions about the ideas presented in this thesis, I would like to thank my colleagues and former colleagues of the Discourse Studies Group, especially Hans Hoeken, Fons Maes, Leonoor Oversteegen, Gisela Redeker, Jan Renkema, Ted Sanders, and Wilbert Spooren. I had valuable consultations with Hans-Georg Bosshardt, Tony Jameson, and Remco Scha. I thank Walter Daelemans for patiently teaching me the basics of LISP, Guillaume Beijers for helping me with statistical issues and Reinier Cozijn for solving hardware problems. Hans Verhulst, Marian Sanders, Pieter Nieuwint, and Marc Vitullo were very cooperative in correcting my English within a very short period. The professional insights of my friends Marleen van Heck, Helma van Heerbeek and Renate de Vré showed their value in making the text fit to be printed.

Part of my way through the labyrinth was through the forest. The daily walks with Til van der Pol, Saskia Schenning, and Maartje Tel provided me with fresh air and fresh thoughts and made it possible to enjoy the seasons changing again and again.

Finally, I would like to thank other friends and relatives, who also supported me during the time I worked on this thesis. I cannot possibly mention them all, but I would like to make an exception for Joost Maas and my mother. Both shared my ups and downs all along the way and still managed to keep unconditional faith in me. Joost not only proved to be a great cook, but he also put things into perspective with a great sense of humour. In addition to many other values, my mother taught me that honesty, respect, and perseverance are essential in life. Therefore I dedicate this book to her.

Els van der Pool

Tilburg, July 1995

Contents

1 Introduction

1.1 Writing as a conceptual process

Many processes in human behaviour are hidden in 'the black box' of our mind. Some of them will probably remain a complete mystery. Others, however, are gradually being unfolded by research. Language production is one of them.

In recent decades different disciplines have contributed to exploring, describing, and explaining (aspects of) language production. Psycholinguists and cognitive psychologists have studied processes of language production in different fashions, on several levels (Garrett 1975, 1980; Butterworth 1980; Bock 1982; Levelt 1989; Emig 1971; Flower & Hayes 1980, 1981; Matsuhashi 1981; Bereiter & Scardamalia 1987). Most research, however, has been focused on spoken language production, probably because it is "one of our dearest occupations" (Levelt 1989: xiii).

Speaking is a complex process as it appeals to many different skills, ranging from articulation, formulation, to conceptualization and communicative interaction. In particular, it is this interaction which distinguishes speaking from writing. In speaking one can immediately interrupt his[1] participant, this is not possible in a written mode.[2] In speaking, the interlocutors provide each other with cues for retrieving information and adjusting it. In writing, however, the writer must be self-sufficient. Since such external incentives are missing, writing is a different kind of task (Bereiter & Scardamalia 1987: 55-56).

This isolated process of language production requires a form of internal reflection. In a conversation the listener may ask for clarification through the use of expressions such as 'what do you mean' or 'I do not understand'. Such reactions on the part of the listener push the speaker to immediately explain his utterance. This kind of external on-line feedback is missing during writing. Writers have to internalize this type of reflection on what they produce. They have to ask themselves 'what do I mean' and 'would the reader understand what I'm writing?'. These reflections may concern the choice of words (Is this word too difficult for the reader, is there a simpler equivalent?), or the construction of sentences (Is this passive construction really needed?). In writing a complete text these reflections have a greater range (which is the best way of ordering information?).

Insight into these conceptual processes of writing on a text level and the development of these processes as people age are still on the frontiers of scientific knowledge, as the processes have only been explored in any detail in the last few decades.[3] Research on the text level has yielded a number of models of the writing process which are diverse in character (see for instance Emig 1971, De Beaugrande 1980; Flower & Hayes 1980, 1981; Bereiter & Scardamalia 1987; Hayes 1992). Research on developmental aspects of the writing process has focused primarily on characterizing types of writing (see, for instance Graves 1975; Pianko 1979; McCutchen & Perfetti 1982; Faigley & Witte 1984; Scinto 1986; Witte & Cherry 1986; Hudson & Shapiro 1991).[4]

In order to study conceptual processes empirically, certain methodological problems must be overcome. It is not possible to gain direct access to the processes hidden in a writer's mind. Thus, one has to gain entry to the human mind indirectly. This entry can be obtained through the products of the conceptual processes. It is assumed that texts contain traces of the conceptual activities (Cooper 1983; Tamor & Bond 1983; Bereiter & Scardamalia 1987: 41). Within the field of research on writing processes, one stream has concentrated on analyzing the products of writers (see Matsuhashi & Quinn 1984; Frederiksen 1986; Witte & Cherry 1986; Witte 1987). On the basis of this observational research, cognitive processes are inferred.

Following this line of empirical research, this study aims to contribute to the exploration and description of the nature and development of the conceptual processes of text production. In addition to the insights already available, this study attempts to contribute to an *explicit* and *systematic specification* of the conceptual processes by analyzing products.[5]

1.2 A text-analytical study of the nature and development of conceptual processes in writing

Using products to infer processes, the question is what traces can be found in texts and what insights do these traces provide into the conceptual activities? Below, two texts are presented as examples. The first text, presented in (1), was written by a 10 year old boy on primary school level. The second text, presented in (2), was written by an adult on university level. These texts, which were originally written in Dutch, were translated into English almost literally; therefore they have a distinctive Dutch flavour (see appendix to chapter 1 for the original Dutch versions).[6] The texts describe a person the writers would like to resemble.

(1) He is a teacher in Heervarebeek. He teaches children to learn like arithmetic Language History Geogaphy. He is round about 40 years old. He wears a suit and a white shirt, and black shoes and looks healthy. Children teaching. On Sundays, Wednesday Afternoons, and Saturday Afternoons cycling Correcting notebooks Walking in the woods. His naame is Uncle Harie. Zometimes he goes to tilburg To the families and window-shopping on Sundays as well.

(2) I would, if forced to make a choice, like to be like Ed Nijpels. He looks quite nice, is well-dressed, and expresses himself very well. Besides, he has undertaken all kinds of activities alongside his rather varied career. It is this variation that appeals to me very much. His "main job" at this moment is being mayor of Breda, but as a compere on TV he holds his own as well! When he relaxes after his busy work, he can be found on the tennis court and the evenings he likes to be with friends or just on his own reading a nice book! Because of his open character he makes friends easily and makes contacts easily. As a mayor he tries to stand close to the people. The age of Ed Nijpels (round about 40) is an age which has many advantages: one has had the chance (and the time!) to make something of one's life. One does not need to be a 'he-man' any more. Summarizing, I must say that in any case it looks as if Ed Nijpels feels comfortable or to put it informally 'he likes who he is' and thàt is something worth striving for!

These two texts differ greatly. The first text contains numerous spelling errors and ill-formed sentences. In reading the first six sentences it is not very clear whom the text is about. In addition, the ending is somewhat abrupt.

The second text is free of spelling errors and grammatically incorrect sentences. After the first sentence, the reader can form specific expectations of what will follow: in all likelihood a description of Ed Nijpels, the man the writer wants to resemble. This text can be qualified as more informative. It is longer and provides more details of the person described. In addition, the text concludes with an explicit marker signalling the end of the description ('Summarizing, I must say...').

What do these observations tell us? From the length of the text, we might deduce that the writer of the first text has less knowledge of the topic than the second writer. In addition, one could remark that there are differences in linguistic knowledge; the first writer's grammar leaves much to be desired, in contrast to the second writer, the first writer does not formulate his sentences adequately. Moreover, the first writer is probably not very much aware of the reader as he 'forgets' to introduce his topic and simply stops at the end, in contrast to the second writer who formulates well and is probably, given the explicit signs, aware of the reader.

Of course such impressionistic observations are inadequate for inferring process aspects. Such an intuitive interpretation is not very objective, as it is rather reader-dependent. Another reader may come up with other characteristics of the texts and, therefore, with another interpretation. Besides, an arbitrary list of remarkable features per text is not acceptable. But how do we examine texts in order to yield a reliable and informative result when no two texts are the same?

The manner in which this problem was approached in this study is depicted in figure 1 (Schilperoord & Van der Pool 1995). This figure also represents the structure of this thesis.

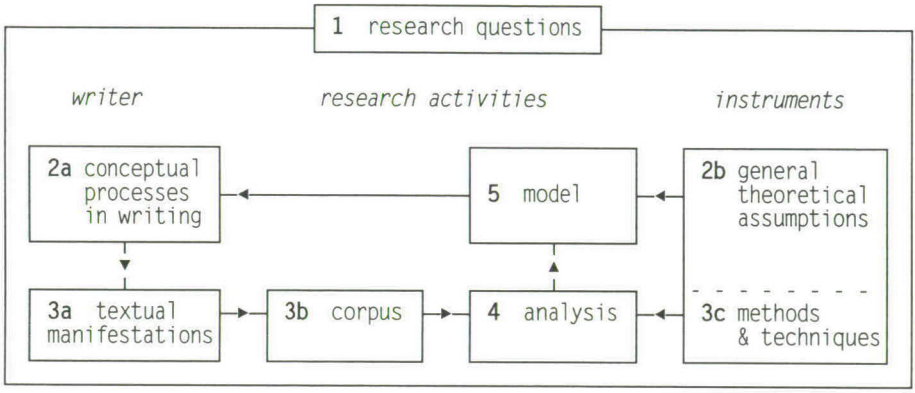


Figure 1. Schematic representation of research activities in studying the nature and development of conceptual processes in writing

Figure 1 displays from left to right: the object of the study (the conceptual processes executed by *the writer*), the activities of the researcher (*research activities*), and the instruments available in studying the writing process (*instruments*).

The study is directed by the *research questions* (see box *1* in figure 1). Our interest is twofold: 'what is the nature of conceptual processes in writing?' and 'how do the conceptual processes develop with age?'. These questions were central in determining how the research activities would take place and which instruments would be used.

Research on the conceptual processes (box 2a) executed by other researchers has yielded general theoretical assumptions (box 2b) about the processes. A psycholinguistic model based on relevant literature concerning the nature and development of conceptual processes

will be presented in *chapter 2*. These views are integrated and presented as a blueprint of the writer. On the basis of these ideas the research questions are specified.

As stated above, the conceptual processes are not directly accessible. Since they leave observable traces in products, these textual manifestations were used in this study as a key to mental activity (box 3a).[7] The substructuring for this choice is presented in *chapter 3*, following a general overview of different types of manifestations and psycholinguistic research methods in general (box 3c).

The manifestations themselves do not, however, inform about the processes; the processes need to be deduced from the text features. Thus, it needs to be figured out which features provide which insights and why. Is the choice of words or syntactic structures informative for revealing conceptual processes? They probably are, but only to a small extent. As conceptual processes primarily concern message generation (the retrieval and organization of information), the structure of a text (in combination with its content) is much more informative. In this study, text structures are therefore used as a means to gain insight into the conceptual aspects.

The specific reasons why text structures provide so much information on conceptual processes will be discussed in *chapter 3*. The chapter is rounded off with a preliminary exploration of the corpus used in this study (box 3b).

As was argued at the beginning of this section, impressionistic observations are not very reliable or informative. An explicit analytical examination of the texts is required in order to transform the observations into usable data (box 4).[8] The analytical instrument used in this study is a procedural method, called PISA, which yields a hierarchical and relational structure of texts. It will be explained in detail in *chapter 4*.

Once the texts have been analyzed, the inductive step to modelling the nature of the processes must be taken (box 5). But how can this be accomplished? This study aims at an explicit and systematic specification of the conceptual processes. In order to achieve that purpose, we induced rules for reconstructing aspects of the conceptual processes executed during writing on the basis of the text analytical results. This inductive step yielded a coherent set of reconstruction rules, called the TRACE model, which made it possible to reconstruct (part of) the conceptual activities executed by the writers of our corpus. This model, which will be discussed in *chapter 5*, is an answer to 'what is the nature of conceptual processes', and a proposal to specify the conceptual processes as represented in the blueprint of the writer.

In order to answer the second research question 'how do the processes develop with age', the TRACE-rules were applied to a corpus of texts in order to characterize the individual performances. Exploration of these characterizations revealed different patterns of rule application. To explore the development with age, we examined the relation between these patterns and the age of the writers. The results of these examinations will be presented in *chapter 6*.[9]

Chapter 7 comprises the conclusion of this study, in which the scope, strength and limitations of the study will be discussed.

Notes to chapter 1

1. With respect to the readability of the text, it was decided to use only 'male' references. Of course, this conventional choice was not meant to restrict the scope of this study. Without wanting to discriminate one of the two genders, it should be remarked that all *he*'s refer to *she*'s as well and what holds for '*him*' also concerns '*her*'.

2. Except for some electronic-mail options, which make on-line dialogues possible.

3. Writing research on the sentence level has a longer history, in particular in relation to development (see Harrell 1957, Hunt 1965, Christensen 1968).

4. See Hillocks 1986 for a comprehensive overview of writing research (from the perspective of teaching).

5. As we are interested in 'natural' language production we wanted to use data which were intrinsic to the production of text, data which would be the result of almost completely natural writing without much interference by the researcher. Other data used in research on conceptual processes are thinking aloud protocols or interviews (see chapter 3).
The reasons for choosing off-line manifestations will be discussed in chapter 3.

6. All texts presented in this dissertation are part of a corpus of descriptive texts (see chapter 3, section 3.4), except for the example in chapter 4 section 4.2; this fragment was constructed.

7. The writing process delivers two types of manifestations: during the process (on-line), for instance, in the form of pauses, and afterwards (off-line), for instance, in the form of texts.

8. See Runkel and McGrath 1972 for this step from observations to data.

9. This part of the study also corresponds to box 5; however, it was directed by a different research question (how do the conceptual processes develop) than the one which corresponds to the results presented in chapter 5.

2 The nature and development of conceptual processes

2.1 Introduction

For the past two decades writing researchers have been paying more and more attention to the study of conceptual processes. The use of new methods such as experimentation and analysis yielded new results which provided a basis for modelling the writing process. Accordingly, writing "has assumed its rightful place within the study of higher cognitive processes" (Kintsch in foreword Bereiter & Scardamalia 1987: ix).

A great deal of research on writing processes has been focused on developmental aspects, especially to the role of strategies, since 'good' and 'poor' writers distinguish themselves in the use of strategies. Strategies provide prominent features to characterize the development of conceptual processes.

In this chapter the theoretical framework is the central issue. General theories, and assumptions related to the conceptual processes and developmental aspects are discussed. This theoretical framework forms the basis for specifying the research questions (see the marked boxes in figure 1).

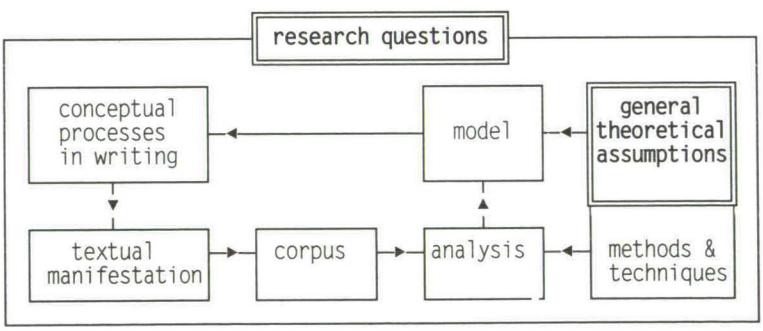


Figure 1. The position of this chapter in the study

Section 2 presents a brief overview of relevant theories on conceptual processes in text production. These theories formed the basis for the 'blueprint of a writer', presented in section 3.[1] This blueprint is a descriptive model in which cognitive and psycholinguistic theories are integrated. The model is used as a framework for exploring, analyzing and describing the

nature and development of conceptual processes in writing.

The development of conceptual processes will be discussed in section 5, and will be preceded by a definition of strategies and procedures in section 4, as these play an important role in the development. The chapter is concluded with the research questions of this study.

2.2 Research on writing processes

The first proposals, based on ancient rhetorics, modelled the writing process as a linear sequence of activities: pre-writing, writing, re-writing. According to this idea, writing mainly consisted of arranging information and choosing a style (e.g. Britton, Burgess, Martin, McLeod & Rosen 1975).

A few researchers, however, were of a different opinion (Emig 1971 among others). They argued that the process is more complex because it is "laminated and recursive" (Emig 1971: 33). Moreover, researchers became aware of the fact that writing takes place in a communicative context. It became clear that a writer -consciously or unconsciously- has to take into account contextual, communicative, linguistic, cognitive and textual constraints (Frederiksen & Dominic 1981).

This shift of attention to non-linearity and communicative aspects contributed to the view that writing is a problem-solving activity (Flower & Hayes 1981; Graves 1981). Writing was conceived as a number of cognitive processes constantly alternating and influencing each other (Witte & Cherry 1986: 121).

These new insights have led to an extensive study of writing processes in relation to skills and text quality (see Breetvelt 1991 for an overview[2]). Descriptive research on the conceptual processes, however, has been given less attention. Cognitive psychologists have developed some over-all models representing insight into the conceptual aspects of writing (Flower & Hayes 1980, 1981; Hayes 1992; Cooper & Matsuhashi 1983; Bereiter & Scardamalia 1987). The two most prominent models in literature, those by Flower and Hayes (1980) and Bereiter and Scardamalia (1987), will be discussed in the next two subsections.

2.2.1 The model of Flower and Hayes

Flower and Hayes (1980, 1981; Hayes & Flower 1980) proposed a model of tasks involved in writing a text. The model is displayed in figure 2.

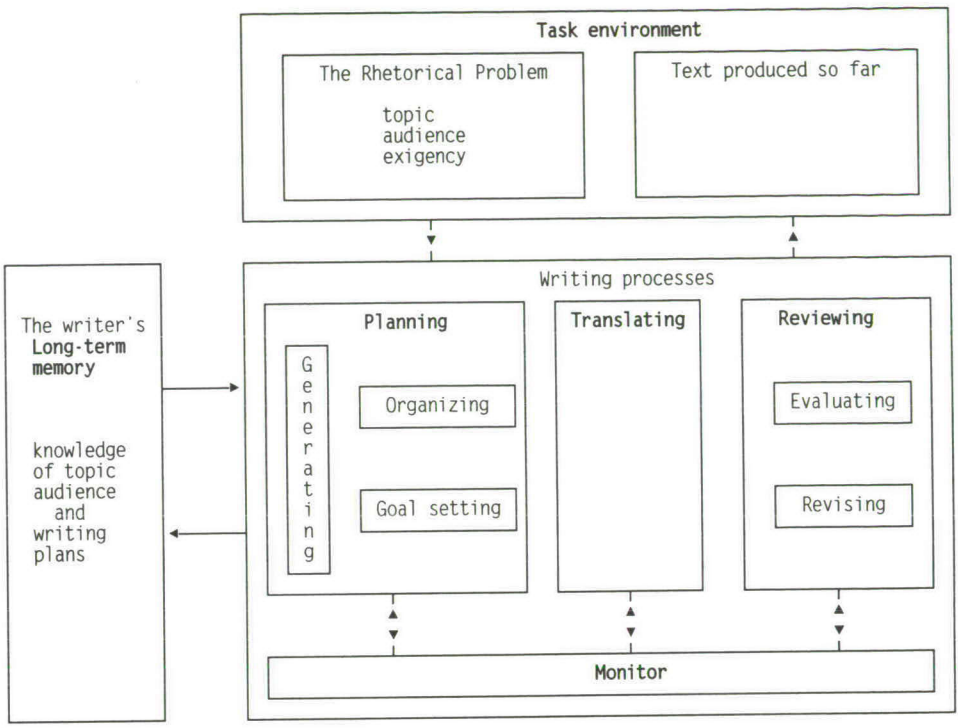


Figure 2. The writing process according to Flower and Hayes (1981)

The writing process consists of planning, translating and reviewing. The process starts with a *planning* phase during which goal setting takes place and ideas are generated and organized. The next step consists of *translating* these ideas into words, phrases and sentences, taking into account the rules of syntax, spelling, etc. Subsequently, the resulting text is reviewed. In *reviewing*, a writer rereads and evaluates the text. These mental processes are checked by a *monitor*, which determines if and when a process takes place and when it is interrupted and restarted.

When planning, translating and reviewing, a writer uses information from the *long-term memory* (which includes knowledge of the topic and audience) and the *task environment*, in which both the rhetorical problem and the text-produced-so-far are stored.

This model has exerted much influence on the research of writing processes, since it has provided the theoretical framework for many studies. But even though the model is very

popular and frequently used, it has also been criticized (Cooper & Holzman 1983, 1985; Van der Geest, Schellens & Van Waes 1992; Vanmaele & Lowyck 1990; Vos 1985), not in the least for its methodological aspects. As Flower and Hayes remark themselves, the empirical basis of their model is rather small. "Although the model was derived through informal analysis of many protocols, it has been tested formally with only one protocol" (Hayes & Flower 1980: 27). In addition, one could dispute the use of the thinking aloud method in order to reveal cognitive processes. In thinking aloud only the goal-oriented processes can be expressed, whereas the more automatized processes remain unexpressed (Cooper & Holzman 1983, 1985).

However, as all research implies methodological choices and as a consequence limitations, this brief discussion of Flower and Hayes' model will be restricted to two major objections related to the content of the model. One involves the distinction between the concepts *planning* and *reviewing*, the other concerns the scope of the model.

Planning and *reviewing* are presented as two separate operations which are the least specified ones (see Hayes & Flower 1980 for a schematic specification of the other processes).[3] This separate presentation suggests that they are completely different (Vanmaele & Lowyck 1990: 216; Cooper & Holzman 1983, 1985). This distinction neglects the similarity in the content of the processes; the processes of planning and reviewing both operate on the content of the message and on rhetorical problems. From a prescriptive point of view, planning and revision are different tasks. Planning is supposed to precede translating, whereas reviewing comes after it.[4] From a descriptive perspective, however, they consist of the same cognitive processes and resources.

Although Flower and Hayes have attempted "to be descriptive rather than prescriptive" (1981: 55), their model resembles a task model rather than a descriptive process model since it represents the ideal way in which the writing task is executed.

The other objection to Flower and Hayes's model is related to the previous one. The scope of the model is restricted to expert ways of writing. The writing process is represented in an ideal form in which much attention is paid to problem-solving activities. It does not represent minimal performances, which occur in novice writing, for instance. This latter way of writing remains unspecified (Vanmaele & Lowyck 1990: 215). As Hayes and Flower state, their model "is a model of competent writers" (1980: 29). They do not pay much attention to novice ways of writing, and leave developmental changes aside.

Hayes (1992) presents the outlines of a new tentative model for text production. He attempts to include cognitive and affective processes as well as the social and material context. Hayes assumes that there are three important cognitive processes in writing:

- language processing, in which mental representations are constructed;
- problem-solving, which consists of operations applied to the mental representations aiming to create new ones;
- text production, in which text is produced on the basis of the mental representations.

The model specifies relations between writing on the one hand and reading and listening on the other. This is best illustrated in the new proposal for the revision process in which reading is incorporated. In addition, there is a brief discussion of two factors that have a special influence on the writing process: planning and motivation.

The proposal only presents the outlines and "tentative ideas" that influence the development of the new model (1992: 247). It is therefore hard to judge whether the objections about the 1980-model are refuted or not.

2.2.2 The models of Bereiter and Scardamalia

Bereiter and Scardamalia (1987) introduced an influential representation of the writing process as well. They present two models: a knowledge-telling model and a knowledge-transforming model. The key distinction between the two models is that the knowledge-telling model depicts writing as a *naturally* acquired capability. The knowledge-transforming model, on the other hand, depicts writing as a *problematic* one. By 'natural' they mean "the kinds of abilities that are most inevitably acquired through ordinary living (including ordinary living in school classrooms)"; by 'problematic' they mean those abilities "that require some special effort to transcend naturally occurring limitations" (1987: 4).[5]

By presenting two models, Bereiter and Scardamalia try to account for both ends of the natural-to-problematic continuum and for interactions, such as to what extent the naturally acquired capabilities are embedded in the more problematic ones.

In writing according to the *knowledge-telling model*, a writer makes maximum use of present competence and skills gained by ordinary social experience. These skills provide an efficient basis for text production (1987: 9).

Writing according to the *knowledge-transforming model* is conceived as a problem-solving task. A writer constantly formulates and solves problems while generating a text. This results in a two-way interaction between developing knowledge and producing text (1987: 12). According to this model, writers "are likely to consider not only changes in text, but also changes in what they want to discuss" (1987: 11).

The model of knowledge-telling is depicted in figure 3.

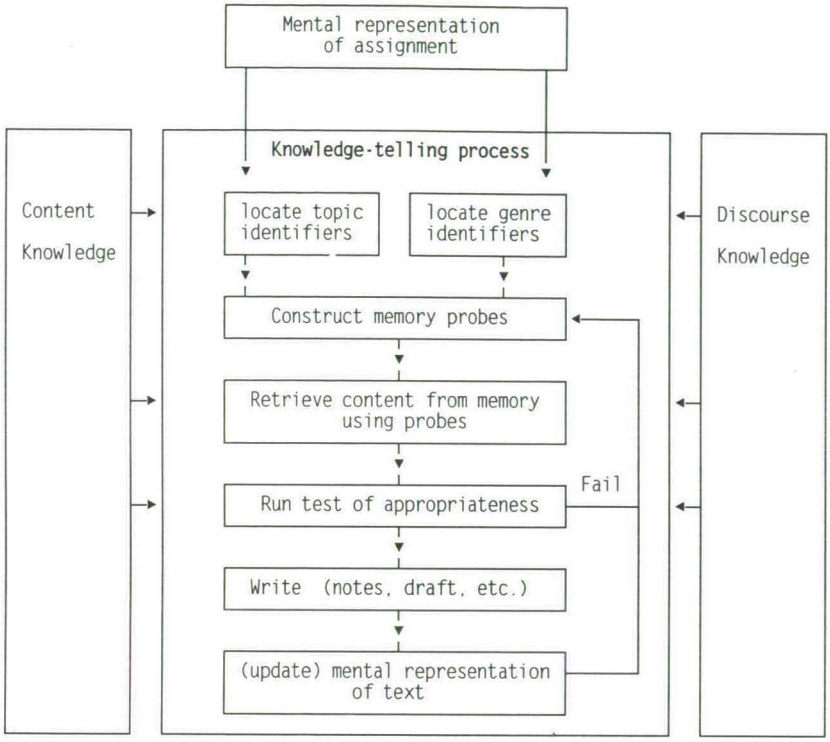


Figure 3. The knowledge-telling model (Bereiter and Scardamalia 1987)

When a writer starts to write, he is guided by a *mental representation of the assignment*. The assignment is analyzed into cues for retrieving information related to the topic (*topic identifiers*) and the genre (*genre identifiers*). These genre identifiers consist of discourse schemas. A discourse schema in turn consists of knowledge of the literary form, the kinds of elements to be included and something about their arrangement. The text-produced-so-far functions as a cue in memory search as well.

In *retrieving information* content and discourse knowledge are used. The retrieved content is tested for its *appropriateness* to genre, to the assignment, etc. If the item passes the test, it is *written down* and the mental representation of the text is updated. A new cycle of knowledge-telling may start (cf. Flower & Hayes's recursiveness of the writing process).

This natural and efficient way of writing can be extended by problem-solving operations involving content and rhetorics. The addition of these operations to the knowledge-telling model is represented by the knowledge-transforming model. This more sophisticated way of writing is depicted in figure 4.[6]

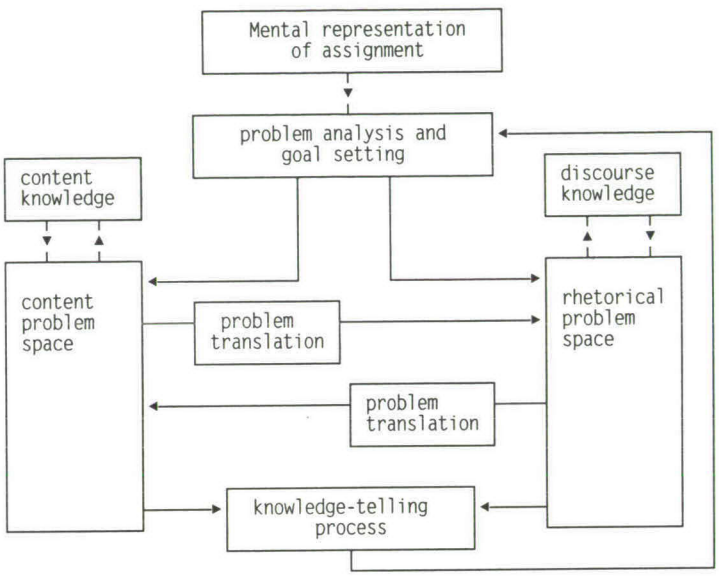


Figure 4. The knowledge-transforming model (Bereiter and Scardamalia 1987)

A writer does not immediately start retrieving content. He starts by making an *analysis of rhetorical and content problems* and by determining communicative *goals*. Rhetorical problems are related to achieving specific goals, such as 'how can I write a comprehensible text for a private investor?'. Content problems are related to belief and knowledge, for instance 'What does the Amsterdam stock exchange mean exactly by limit orders ?'. These problems are dealt with in two different problem spaces which interact with each other. In the example of writing a comprehensible text for a private investor, the concept of 'limit orders' needs to be explained.

The results of these reasoning processes serve as input for the *knowledge-telling component*. This component generates textual output, which is reflected upon by analyzing problems and checking goal setting, and which may open a new cycle of transforming knowledge. This is a cyclic way of writing because it provides feedback to the knowledge-telling process.

The distinctive features of the knowledge-transforming model consist of formulating and solving content and rhetorical problems. This problem-solving activity takes place in a way which allows a two-way interaction between continuously developing knowledge as well as continuously developing text (1987: 12). A constant reflection on the text and knowledge provides a writer with new insights. This reasoning not only takes place before information is written down, but also afterwards when the product of writing is evaluated by problem analysis and goal checking.

Unlike Flower and Hayes, Bereiter and Scardamalia not only characterize an 'expert'-like way of writing (knowledge-transforming), but also a 'novice'-like one (knowledge-telling). Bereiter and Scardamalia used two discrete models because they wanted to emphasize the distinctions between design concepts (1987: 29). In that respect they certainly succeeded, for each model corresponds to a different *strategy* of writing.

However, these two different ways of writing are not necessarily linked to a specific *type* of writer. Writers who are capable of producing a text according to the knowledge-transforming model, are not always inclined to do so. It might be more efficient to use the knowledge-telling way under certain communicative circumstances. If a lawyer, for instance, has to write a routine letter to a client, knowledge-telling will do (Schilperoord 1995). The knowledge-transforming model represents the capacities of experts, but these are not necessarily used in all communicative situations.

Furthermore, writing according to the knowledge-transforming model does not always involve all aspects (Bereiter & Scardamalia 1987: 29). Some tasks require different forms of reasoning than others. If one was to write a historical overview of the Amsterdam stock exchange, this will entail fewer rhetorical problems than writing an argumentative text in which foreign investors have to be persuaded to do business in Amsterdam. In writing a historical overview the main rhetorical problem is related to the writer's awareness of his audience: he will, for instance, have to prevent the readers from getting bored. In generating a persuasive text, on the other hand, a writer may have to convince the reader to come to Amsterdam. Since such argumentation can assume several forms, the writer has to consider which form he thinks will be most effective.

In order to get a complete view of developmental aspects, it is more interesting to characterize differences between writers gradually than in terms of two extremes. *One flexible* model is therefore preferred.

Because the knowledge-transforming model includes a 'minimal' as well as an 'extended' way of writing, this model can serve as a basis for representing developmental changes.
We regard the knowledge-transforming model as a maximal option, which is not necessarily executed completely. This interpretation allows for a gradual development from immature to mature writing in one model, and it enables a characterization of strategy switches within mature writing.

2.3 A blueprint of a writer

This study requires a framework which represents the content of the processes and which allows for a representation of the gradual development. Thus, the above-mentioned models and ideas have been integrated in one comprehensive *blueprint of a writer*. The model aims at a descriptive representation of the processes in writing, especially the conceptual ones.

The blueprint is not only influenced by the writing models already discussed, but also by models of spoken language production, especially by Levelt (1989) and Dijkstra and Kempen (1993).[7] The models for speech production are more differentiated, as they have a greater experimental and empirical basis with respect to the 'lower', more basic processes, especially the linguistic ones.

We consider these models of oral production of great relevance for the writing model. In spite of the differences between spoken and written language production, both modes show clear structural resemblances. There are obvious similarities between their components and products; thus, for instance, both use similar units on the grammatical and phonological level (Allport et al. 1987).[8] Because of these resemblances we generalized the essential characteristics of the models on these levels.

2.3.1 The outlines

The blueprint of a writer is depicted in figure 5.

On the left hand side of figure 5 (top to bottom) components of the writing process are represented: **Inventing** a message, **Encoding** it into natural language and **Executing** motor activities to write it down. These production components are of three different types: *inventing* is a conceptual type of process, *encoding* a linguistic one, and *executing* a physical one (Dijkstra & Kempen 1993).

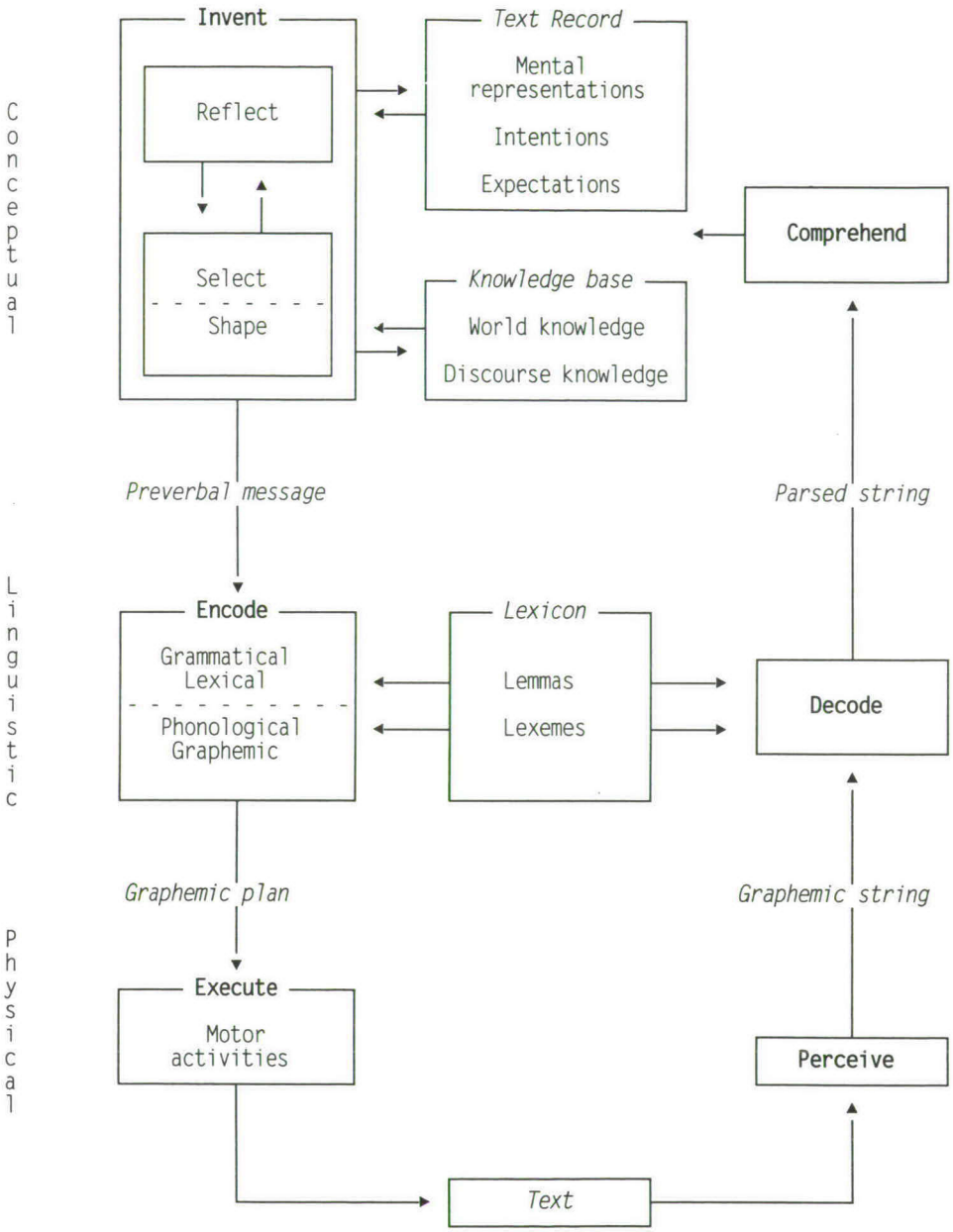


Figure 5. A blueprint of a writer

Text production starts with *inventing* information. This results in a preverbal message, which is a semantic representation of the message to be expressed (Levelt 1989: 73). The preverbal message is subsequently *encoded* in natural language by assigning linguistic features, such as grammatical form and phonological aspects (Bock 1982; Garrett 1980; Kempen & Hoenkamp 1987; Meyer 1988). This formulating process provides a graphemic plan, consisting of a (rhythmic) syllabification of a string of segments (cf. 'phonetic plan' in Levelt 1989: 284). Finally, by means of hand movements with a pen or on a keyboard the information is written down (Thomassen & Van Galen 1992). This process as a whole results in a text.[9]

Text Record

When writing, an author not only creates a physical version of a text, but a mental representation of the text(-produced-so-far) as well.[10] This mental representation is kept in a text record, being a temporary store of information about the text and of the context, "currently available or accessible" to the writer (cf. 'discourse record' in Levelt 1989: 111).

This text record also comprises the writer's intentions and expectations.[11] They have a guiding function in producing text as they influence the retrieval and the organization of information. If a writer wants to persuade a reader, for instance, he will select different information than if his aim is to amuse (Matsuhashi 1981). Besides, in persuading a reader, a writer will select those arguments that support his opinion best. The same holds for the organization of the information. He will present the strongest arguments in a more prominent position than those he assumes are less convincing.

Apart from intentions a writer also has expectations, for instance about the knowledge, background, and intentions of the potential reader.

Knowledge base

The conceptual component makes use of a knowledge base which includes world and discourse knowledge. Both include declarative as well as procedural knowledge.

Declarative knowledge can be compared with encyclopedic knowledge. It is related to facts (propositional knowledge) and the discourse situation (cf. 'situational knowledge' in Levelt 1989: 10). Examples of declarative world knowledge would for instance be 'The mayor of Breda is Ed Nijpels' or 'Stefan Pettersson scores many goals'. Declarative discourse knowledge takes the form of discourse schemas (as Bereiter and Scardamalia (1987) use in their models). A discourse schema consists of knowledge of a selected literary form (such as narrative or argument), which specifies the kinds of elements to be included in the discourse and something about their arrangement.

Procedural knowledge takes the form of if-then rules, which are applied to declarative knowledge (Levelt 1989: 72). Examples would be 'If you are looking for a good football player, take one from Ajax', and 'If you have to write a business letter, start with a salutation, followed by an introduction, a central message and a closing'.

Within the conceptual component, reasoning processes take place on the basis of the available knowledge. These processes yield new insights for the language user, because the conceptual component can alter knowledge on the basis of reasoning. As a consequence, there is a two way interaction between the knowledge base and the conceptual component. This is depicted in figure 5 by two arrows symbolizing the interaction: one arrow from the knowledge base to the invent-component, and one going the other way.[12]

Lexicon

The linguistic component uses the lexicon. For every element in the lexicon different kinds of information are available (Dijkstra and Kempen 1993). The 'lemma'-part contains syntactic and conceptual information, for instance, 'to give' is a transitive verb with three cases (agens, object and indirect object) which can take the meaning of 'to pass', 'to administer' or 'to supply with'. The 'lexeme'-part contains morphological and phonological features, for instance, it is a 'strong verb', the past tense is characterized by vowel gradation ('gave'), and the stress in 'given' is on the first syllable,.

In contrast to world knowledge, the lexicon is less dynamic. The use of the lexicon is more of a one-way activity. This is symbolized by the single arrow from the lexicon to the encoding and decoding components.

Perceive, Decode and Comprehend: the perceptual loop

Writing generally consists of more operations than *Inventing*, *Encoding* and *Executing*. Once a text has been produced, a writer may read his own text in order to revise elements. This editing of one's own text is depicted as a '*perceptual loop*' consisting of *Perceiving*, *Decoding*, and *Comprehending* (see figure 5).

During or after producing a text a writer may wonder for instance whether his punctuation is right, whether his spelling is correct, whether his syntax and morphology are correct, whether what he is writing is up to social standards, whether it is the way he wants to say it. If the evaluation is negative, he may decide to adapt the output.

This evaluation of the (in-between) output results from a complicated interplay of comprehension and production processes, in which a writer has to execute a number of

operations in order to adapt the output (Levelt 1983: 45-47). In short, he first has to detect errors or any other inappropriate elements, he subsequently has to consider whether they should be corrected or not, and if so he has to interrupt the production, and finally he has to adapt the output in an adequate way (see Dirksen 1990 chapter 3).[13] All these operations take place within a fraction of a second.

There are two entirely different ways of looking at this revision of output (Levelt 1983: 46). The first approach assumes that a language producer has direct access to particular components of the process. By evaluating the output of the components, the language producer is alarmed if the evaluation is negative. He may then decide to stop in order to reconsider the production. Levelt calls this approach 'the production theory of monitoring' (1983: 46). This theory implies that every error is detected almost immediately (Levelt 1989: 477).

In the other approach the producer is assumed to have no access to the processes, but only to their final results. Comparing the output with the original intention, the language producer can detect and repair troublesome output.[14] This view is called the 'perceptual theory of monitoring' by Levelt (1983: 46).

The *perceptual loop* is an economical theory of revising output as it avoids unnecessary 'doubling' of devices. From an efficiency principle of production, it is likely that the capabilities for editing other people's products are also used in editing one's own product. This approach does not require any additional revision devices.

In running the perceptual loop, the text is processed by the normal language system (Levelt 1989: 469). This is represented in the model by the three components: *Perceive*, *Decode*, and *Comprehend* (see figure 5).

In reading one's own text a series of characters are *perceived* by means of sensory activities of the eyes. This results in a *graphemic string*. *Decode* processes reconstruct from this string the lexical content and the syntactic form of the clause-like elements. A *parsed string* is the result. Parsed strings are connected to each other and enriched with world knowledge in the *comprehend* component. The result is a mental representation stored in the text record (Noordman & Maes 1993).

Running the perceptual loop in order to revise one's own text differs from reading someone else's text, because the revising writer has access to different information.

Reading a text, one builds up a representation of what is perceived. If someone else's text is read - or if it is a long time ago that one's own text was written - all words and phrases have to be decoded and comprehended in order to create a mental representation. Although there may be strong expectations, the text can be considered as 'new'.

Reading one's own text in order to revise it (just after it has been written down), one also tries to build up a mental representation of the text. However, there is already a representation of the text, one which was created previously during the writing. It is this representation that may interfere with the revision of the product. What is perceived might not be the pure text, but a mixture of the written text and the previously created mental representation. The presence of this representation can speed up reading times in processing one's own text. On the other hand, it may also complicate the detection of errors (Levelt 1989: 463). The representation may interfere with the reading of the written text. What is perceived is not only the text but also the mental representation built during the writing.

Detection of malfunctions or discrepancies in the writing process therefore not only depends on the writing skill and the difficulty of the problem. More important is the writer's ability to momentarily disregard the interpretations based on his writing intentions. In revising his own text, a writer should "inhibit a tendency to use 'privileged' knowledge" (Bartlett 1982: 350). The notion of privileged knowledge, of course, only refers to writers' knowledge of their own intentions. Obviously, writers will not have such information about texts composed by others and consequently will not be faced with the problem of inhibiting certain interpretive activities when reading other people's texts (Bartlett 1982: 350).[15]

Because of this privileged knowledge, it is harder to take the perspective of an 'unprejudiced' reader in revising one's own text. Not all problems are noticed or given equal attention, because of the interference of what is meant (intentions) and what has actually been realized. As Murray states, writers must achieve a detachment from their work that allows them to see what is on the page, not what they hope will be on the page (1978: 95). Levelt calls the phenomenon of not giving equal attention to malfunctions 'selective attention' (1989: 467).

How the perceptual loop operates exactly is not the issue here, and will therefore not be discussed in any more detail.[16]

2.3.2 Conceptual processes in writing

Psycholinguistic research has mainly concentrated on linguistic and physical processes (encoding and motor activities). This has yielded an amount of valuable results on the basis of which these components can be specified. The conceptual component, however, is less specified. Research on conceptual processes in language production has been conducted mostly since the beginning of the eighties. Prominent models of written text production are those by Bereiter and Scardamalia (1987) and Flower and Hayes (1980, 1981); for speech production those by Levelt (1989) and Dijkstra and Kempen (1993). Combining and integrating their ideas about conceptual processes has led to the specification depicted in Figure 6.

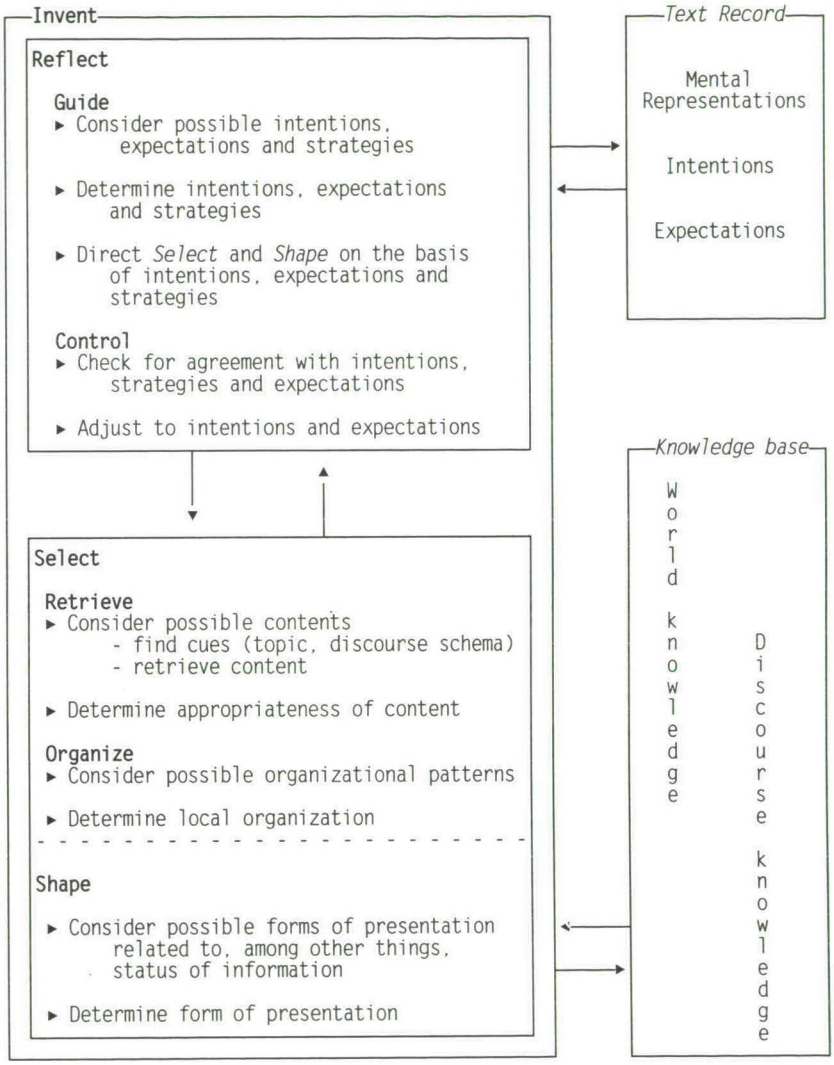


Figure 6. Specification of the Invent component

In *inventing* a message, a writer reflects on intentions and expectations (*Reflect*). In addition, a writer retrieves and organizes information (*Select*) and chooses a form to present it in units (*Shape*). These three operations within *Invent* correspond to subprocesses of the conceptual components in models proposed by other authors (see table 1).[17]

Table 1. Correspondence between operations within the Invent-component and subprocesses in existing models

Authors	Reflect	Select	Shape
Flower & Hayes (1980)	goal setting, evaluating, rhetorical problem	generating, organizing	--
Bereiter & Scardamalia (1987)	conceptual planning,i.e. problem and goal setting in knowledge-transforming	content generation in knowledge-telling	--
Levelt (1989)	macroplanning, i.e. specifying communicative goals	macroplanning, i.e. retrieving information	microplanning
Dijkstra & Kempen (1993)	forming goals	forming ideas	microplanning

The *Reflect*-operation corresponds to Flower and Hayes's 'goal setting' and 'evaluating' on the basis of a 'rhetorical problem' (1980, 1981). This is similar to Bereiter and Scardamalia's 'conceptual planning', being 'problem and goal setting' represented in the knowledge-transforming model (1987). In Levelt's model *Reflect* corresponds to 'macroplanning', which involves specifying communicative goals. Dijkstra and Kempen take over the view that *Reflect* corresponds to 'forming goals' (1993).

The *Select*-operation corresponds to Flower and Hayes's components of 'generating' and 'organizing' and to Bereiter and Scardamalia's 'content generation' in knowledge-telling, consisting of locating identifiers, constructing memory probes, retrieving information and testing its appropriateness. *Select* corresponds to Levelt's 'macroplanning', which is 'retrieving information', and fits in with what Dijkstra and Kempen call 'forming ideas'.

The *Shape*-operation of the conceptual component corresponds to Levelt's 'microplanning', which is taken over by Dijkstra and Kempen. The other models do not distinguish this function.

These *Invent*-operations are *recursive* and *iterative*; they alternate and may be repeated. Thus, Reflect may interrupt Select and/or Shape to evaluate their (in-between) output against the light of intentions and expectations, and restart them again (cf. Hayes & Flower 1980; Bereiter & Scardamalia 1987).

Reflect

During *Reflect* a writer examines possible communicative intentions, strategies and expectations, such as 'what do I want to achieve with my text?', 'who will read the text?'. He determines intentions on the basis of world and discourse knowledge, his own beliefs and the contextual situation. These intentions and expectations *guide* the Select and Shape operations (Bereiter & Scardamalia 1987: 193; Levelt 1989: 110). This guiding function mostly precedes Executing and is therefore often referred to as 'planning'.[18]

Intentions and expectations are not only used to guide Select and Shape, but also to *control* them. The output of Select and Shape is checked for agreement with the specified intentions, strategies and expectations. If a writer thinks that his intentions are inadequately realized and/or the output does not come up to the expectations, he will adjust the output. This is realized by having Select and Shape operate again.[19] If the evaluate operation takes place after the text has been written down, it is generally called 'revision'.[20] Summarizing, Reflect supervises Select and Shape before, during and after operating.

Select

Generating information consists of determining content and structure (cf. 'macroplanning' in Levelt 1989: 123 ff.). We call this operation *Select*. The term Select should be interpreted broadly, as inorderly retrieving as well as organizing information.

In order to be able to retrieve information it has to be activated first. This activation takes place by means of cues related to the topic, for instance the text produced so far, and a discourse schema. A discourse schema specifies content elements of a certain text genre (e.g. narrative or argumentative) and their organisation in the text (Bereiter & Scardamalia 1987: 7).

After information has been activated and retrieved, it is tested for its appropriateness for being incorporated in the message (cf. 'test of appropriateness' in Bereiter & Scardamalia 1987: 9).

Information that passes the test is subsequently organized locally. Possible organization patterns are examined and directed by intentions and expectations and an organization format for the message is determined (e.g. 'linear order' in Levelt 1989: 138).

Shape

Before the selected information can be expressed verbally, form features will have to be determined (cf. microplanning in Levelt 1989: 144 ff.). What has to be determined is, among other things, what information is in focus, what information is given or new and which elements have to be topicalized. The choices for these form aspects are context-dependent.

These decisions are made on the basis of previously Selected and Shaped information and on the basis of the intentions and expectations.

These operations will be illustrated by an example of how the *Invent*-component might operate.[21] Suppose someone is going to write an application letter, how does *Invent* operate? Table 2 presents an example of the part of the letter in which the writer argues that he is qualified for the job and that hopefully he will be selected for a personal interview.
The example will be discussed in the light of the operations depicted in figure 6.

In **Reflecting** an applicant has to determine his intentions ('I want to convince them that they should invite me for an interview'), and find out the expectation ('the reader wants to know why I am qualified'). In order to write an application letter he will have to be more specific.[22] If he wants to be invited for an interview, possible writing goals are 'give a self-assured presentation' (IIa), 'show your competence' (IIb), 'emphasize your positive characteristics' (IIc). If he expects that a reader will want to know whether he is qualified or not, he should mention relevant salient aspects (IVa) and give an argumentation (IVb). These specified intentions and expectations guide and evaluate the Select and Shape operations.

Selecting consists of retrieval and organization of information. Retrieval of information includes testing its appropriateness by matching it to, for instance, the intentions and expectations. An applicant has to determine what information is appropriate in order to convince the reader of his qualities. If a writer aims at showing his best side, it will not be very wise to include less positive characteristics, such as being too old (Vd). The applicant may even decide to conceal the information about his limited experience.

In addition, the writer has to determine a suitable order in which to present the retrieved information. An option to organize the information on a local level is to substructure his claim by telling that he is an appropriate candidate (VIa). He may also decide to do it the other way round: present arguments first and conclude with the statement that he is the right person for the job (VIb). The writer might prefer the first order, the one with the most direct approach (VIa) as he intended to give a confident presentation (intention IIa) and to anticipate on what the reader wants to know (expectation IVb).

Table 2. An example of Invent operations in writing an argumentative part of an application letter

Reflect: *What are my communicative intentions?*
Consider possible communicative intentions
Ia I want to inform them about my qualifications
Ib I want to convince them that they should invite me for an interview
Ic I want to show them that I am really trying to find a job
Determine relevant intention(s): Ia and Ib are most important

How can I achieve that with my letter?
Specify intention:
IIa give them a self-assured presentation
IIb show competence
IIc emphasize positive characteristics
IId take advantage of their needs

What are my expectations?
Consider possible expectations
IIIa the reader does not know me personally
IIIb the reader wants to know why I am qualified
Determine the relevant expectation(s): Both IIIa and IIIb are likely to be the case

What are the consequences of this for my letter?
Specify expectation:
IVa the reader looks for relevant salient aspects
IVb the reader wants an argumentation

Select: *What information do I retrieve?*
Consider possible contents
Va required knowledge
Vb studies abroad
Vc some experience, not in all areas
Vd age: older than desired
Determine content (on the basis of intentions IIb and IIc, expectation IVa):
Va, Vb are preferred; only mention relevant experience of Vc; do not use Vd

How do I organize the information?
Consider possible organizational patterns
VIa substructure (appropriate FOR arguments x, y, z)
VIb conclude (arguments x, y, z, SO appropriate)
Determine organization (on the basis of intention Ia, and expectation IVb):
VIa is preferred; begin with a claim followed by arguments

Shape: *How do I present the information?*
Consider possible styles of presentation
VIIa topicalize a referent or not
VIIb determine status of information (given-new)
Determine Shape (on the basis of text constraints, intentions and expectations):
Choose ... (this differs per segment)

In **Shaping** the writer determines a form of presenting the information in small units. He has to consider, among other things, the accessibility status, and the assignment of the topic role of referents.
The applicant may, for instance, choose to mark the desired experience as topic, as is the case in (1a). However, taking into account writing goal Ib, option (1b) should be preferred.

(1a) The desired experience has been acquired in several jobs.
(1b) I have acquired the experience desired in several jobs.

With respect to the accessibility status of the information, the writer may choose (2a) if the experience is not mentioned earlier in the text. If it is mentioned in the sentence immediately preceding it, (2b) is more suitable.

(2a) I have extensive experience designing and presenting language courses for adults.
(2b) I have experience presenting language courses.

These examples show that in shaping it is not only the intentions and expectations that guide the process; constraints imposed by the text written so far also exert their influence.

The conceptual processes described above are intentional, i.e. the writer wants to achieve a purpose following a certain path. If the writer wants to achieve his intention, strategies form an essential constituent within the conceptual processes. A writer's conceptual skill is determined by the available range of strategies, procedures and the related meta-cognition (cf. Flower's 'self-awareness' 1979: 45).[23] As strategies take up such a prominent role within conceptual processes and development, we need to take a closer look at these concepts.

2.3.3 Strategies and procedures: a definition

A definition closely related to the general conception (e.g. Van Dijk & Kintsch, 1983: 62-68) runs as follows: a strategy is a global depiction of the procedure(s) used to realize a certain intention (see (3)).

(3) Intention → What do you want to achieve? (problem)
Strategy → How can you realise that? (solution)
Procedure → How do you operate precisely? (action)

This means that a strategy consists of a scheme of actions set up to solve a problem.[24] As is shown by (3), a strategy is inextricably bound up with a motivated intention and one or

more specific actions, but it can never be equated with either of these.

The relations between intention and strategy on the one hand and between strategy and procedure on the other are only instrumental; in general a strategy cannot be deduced from the intention, nor can the procedure be deduced from the strategy. An intention can almost always be reached by several strategies, and each of these strategies by several procedures. See for instance (4): there are different strategies to teach tax officers how to write comprehensible texts, and in addition every strategy offers several procedures.

(4)	Intention	Teach tax officers how to write comprehensible texts
	Strategy 1	Show them how this is done
	Procedure 1a	Give a lecture from a theoretical perspective
	Procedure 1b	Give an example from a practical perspective
	Strategy 2	Have them discover what is wrong in their own texts
	Procedure 2a	tell them (monologue)
	Procedure 2b	ask them (interactive)
	Strategy 3	Offer solutions to possible writing problems
	Procedure 3a	present a fixed overall structure
	Procedure 3b	present alternatives to official formulations

As (4) shows, a strategy is part of a network; there may be more strategies to achieve one intention. The above example (4) is a network dominated by one intention and elaborated by several strategies each having their own procedures. The strategies can be used separately or in combination with each other. What they have in common is that they are all possible solutions to one and the same problem 'teaching tax officers how to ...'. However, strategy-networks may also invoke new intentions. See for instance (5) for an example of such an embedded network.

(5)	Intention	I want to know what the text is about globally
	Strategy 1	Look for theme indications
	Procedure 1a	Search for a title, subtitles and illustration (Bimmel & Westhoff 1993)
	Procedure 1b	Search for topic sentences and concluding sentences (Onrust, Verhagen & Doeve 1993)
	Intention-1 deduced from 1b	I want to know what the topic sentences are
	Strategy 1b1	Search for them in the usual places
	Procedure 1b1a	Look at the beginning of each section (MacDonald 1983)
	Intention-2 deduced from 1b	I want to know what the concluding sentences are
	Strategy 1b2	Search for them in the usual places
	Procedure 1b2a	Look at the closing sentence of each section (MacDonald 1983)

Strategy 1 in (5) exemplifies that someone can use a simple procedure (1a) or a more sophisticated one (1b); the difference between 1a and 1b being that the elements of procedure 1a are presented in the text explicitly whereas those of 1b are not. This latter procedure therefore requires activation of sub-goals and accompanying strategies. Both are 'strategic' in their own way, depending on contextual factors (how much time does a reader have, what is the expected importance of the text, etc.).

A problem arises when more than one strategy is available to achieve the same goal. One has to consider carefully which strategy to choose. This is done by relating the intention to the context. This context often determines the restrictions to the strategies and actions. For instance, if you are certain that texts written by tax officers will only improve if the writers' attitude changes, strategy 1 in (4) will not have much effect. So, in selecting a strategy one has to take into account the intention and the context. In other words, strategic working presupposes tactics.

Tactics require a cost-benefit analysis of the intention. The intention is evaluated in the light of an optimum: how directly, well, quickly, economically do you want to reach your goal (Van Dijk & Kintsch 1983). For instance, strategy 2 in (4) is suitable if one wants to persuade the officers of the necessity of revising their texts; strategy 3 is suitable if one wants to motivate the officers and offer them something that is useful to them after the course is finished.

2.4 Development of conceptual processes

The description of the conceptual writing component given so far represents an expert-like performance, guided by intentions and strategies. This is obviously not the way novices operate. The way novices become experts is characterized by the use of strategies (see Faigley, Cherry, Jolliffe & Skinner 1985: 168). The differences in writing between novices and experts in terms of strategies have been the object of many studies, in which the products and behaviours of both groups differing in writing expertise are compared.[25]

Pianko (1979), for instance, distinguishes between 'reacting writing' and 'reflecting writing' to point out differences between writers. Bereiter and Scardamalia, of course, distinguish a knowledge-telling and knowledge-transforming strategy. The presence (or absence) of planning in writing is one of the reasons why Scardamalia and Bereiter designed two models to describe the writing process. Poor writers use planning to create (pre-)text instead of creating higher goals (e.g. Witte 1987). Their planning is much more of a general

repetition for the final text (Bereiter & Scardamalia 1987: 21).

What is found for planning also holds for revising texts. 'Good' writers differ from the 'poor' ones with respect to the level on which they revise. Whereas 'poor' writers mostly revise only on the surface level of the text (spelling and punctuation) and on word and sentence level, 'good' writers also revise the content of their text, e.g. by revising on paragraph or text level (see Breetvelt 1991 for an overview).

The distinction between novice and expert can be specified in terms of Reflect, Select, and Shape. In table 3 the most important results are summarized. The proficiency level varies on a scale from novice to expert.[26]

Table 3. The main differences between novice and expert writers

	Novice	Expert
Reflect *guide*	use literal assignment as intention *task execution* *shallow processing*	determine and specify intention themselves *problem solving* *deep processing*
Select *retrieve*	depend on retrieval cues only *knowledge-telling*	add rhetorical criteria supplied by Reflect *knowledge-transforming*
Select *organize*	present info in order of retrieval *writer based prose*	determine order with respect to intention *reader based prose*
Shape	short sentences, topic as main reference for given information, additive connectors *list-like patterns*	extended more complex sentences, variation in reference, more causal connectors *zig-zag patterns*
Reflect *control*	check content for topic relatedness *add/delete*	check for agreement with intentions *adjust/reshape*

Novice and expert writers differ in the extent to which they are able to Reflect on the assignment. Novices consider writing as a *task executing* activity, as a consequence of which they reflect in a *shallow* or less profound way. A shallow way of operating is characterized by a minimal use of knowledge and a minimal specification of intentions (cf. Just & Carpenter's shallow text processing 1987).[27] Experts are able to conceive writing as a *problem-solving* activity and reflect more intensively on the assignment and the implied constraints. Therefore they are able to reflect *deeply*, as they can make (optimal) use of the accessible knowledge and can lean on more specified intentions in generating a message.

As a result experts place more specific demands on content selection. This deeper reflection takes place in Reflect as well as in Select-operations; in both processes they not only take the assignment into account but also (implicit) constraints related to the assignment.

In addition, the selected information is not just written down in the retrieved order. It is organized with respect to rhetorical goals. This results in a more cyclical way of working, which may lead to new insights and transformation of knowledge. (cf. the knowledge-transforming strategy; Bereiter & Scardamalia 1987). This way of working yields *reader based prose*, in contrast to the novice's way of production. The novices take the assignment literally, select information on the basis of topic and genre cues only, and present the selected information in order of retrieval. As a consequence, novices only use (*tell*) the available knowledge which results in *writer based prose* (Flower 1979).[28]

In presenting their information, novices use short sentences whereas experts use more extended complex sentences (Hunt 1965, 1983; Christensen 1968; Schuurs 1990). The informational status (given-new distribution) is almost the same in every sentence the novices write. The given information is continually the topic (McCuthen & Perfetti 1982). Besides, they link their sentences by using mainly additive connectors, repetition and synonymy (Rentel & King 1983: 164). This yields rather list-like patterns (McCutchen & Perfetti 1982: 120; Daneš 1974). The segments are related to each other mainly through repetition of the topic of the text. The topic is almost the only given tie that is used to elaborate the text. The experts, in contrast, have more varied given-new distributions. In addition to the topic, they also use other given elements mentioned to tie information, and they also use causal connectors more often. All together this yields a more varied pattern (cf. what McCutchen & Perfetti call a 'zig-zag pattern' 1982).

Novices hardly control their texts. If they do, the number of revisions is very small and they are executed in a minimal way: they hardly revise on a conceptual level. Their revisions only consist of a minimal check for topic relatedness on the basis of which they add or delete

elements.[29] Most revisions take place on word- and sentence level (Faigley 1981).

Experts, on the other hand, revise more extensively. In their revision they also try to realize rhetorical goals (Van der Geest, 1991). They not only add or delete elements but also permute and reshape them.

Summarizing, novices are characterized by a minimal operation of Reflect. They concentrate on Select and Shape (cf. knowledge-telling). Experts, on the other hand, can be characterized by a more extended operation of Reflect, which directs Select and Shape more and more (cf. knowledge-transforming). This is congruent with Bereiter and Scardamalia's proposal of a *developmental scenario*: what happens in writing development is the increasing gradual differentiation of Reflect from Select and Shape. Roughly speaking the scenario consists of three phases (1987: 211)[30]:

- "at first all children's conscious attention is involved in the immediate written expression." In our terms this means that *encoding* the message into a linguistic format dominates the message generation.
- subsequently "over the course of childhood and early adolescence, thought becomes sufficiently detached from immediate expression that the young writer can generate text content in abbreviated forms and mentally manipulate it ...". In our terms this means that these novices are mainly concerned with *Select*-operations.
- thirdly, "not until later adolescence we see that ... text organization, intentions, problems, strategies and the like are clearly represented and capable of being operated upon, rather than remaining implicit and in the background". In our terms this means that skilled writers execute *Reflect*-operations more and more.

2.5 Research questions

The aim of this study is to get insight into the nature and development of conceptual processes in writing. As was stated above, strategies and procedures play a crucial role in this. They are global depictions of the actions which need to be executed to solve conceptual problems.

Writing a text can be regarded as continuously solving problems of different kinds. The writer should determine his intentions and expectations[31], he is confronted with, among other things, retrieval and organizational problems, and having invented this message, he has to consider a suitable form of presentation.

These problems imply that Reflect and Select operations are the essence of the conceptual process. They result in a message that subsequently needs to be Shaped in order to be linguistically encoded. Thus, Shape works on the outcome of Reflect- and Select- operations. This study is restricted to Reflect- and Select-operations, as they are the most basic in message generation.

In Reflecting and Selecting, writers (consciously or unconsciously) apply strategies and procedures. The question is what strategies and procedures they use and, in addition, how these strategies and procedures develop with age.

This could imply, however, that it is known what strategies and procedures *exactly* a writer uses in producing text. As, to the best of our knowledge, this is not the case, it should be sorted out first. An empirical way to gain insight into these operations is by reconstructing them. In order to gain insight into the exact operations writers use, this study focuses on the behaviour of individual writers.[32] The aim of this study is not to test predetermined categories of writers, but to explore and describe general patterns in their conceptual processes.

As a consequence, there are two central issues in this study, a methodological one and an empirical one, both including two subquestions:

1. How can Reflect- and Select-operations be studied?
   a. What type of instrument and kind of manifestations have to be used?
      This question will be answered in chapter 3.
   b. What specific form of the instrument is required?
      This issue will be discussed in chapter 4.

2. How can Reflect- and Select-operations be specified in terms of strategies and procedures?
   a. What is the nature of Reflect- and Select-operations?
      This result will be presented in chapter 5.
   b. How do Reflect and Select-operations develop with age?
      This question will be answered in chapter 6.

Chapter 7 presents the conclusions of the study.

Notes

1. An integrated discussion of the model presented in this chapter and the one in chapter 3 together with a discussion about the development of processes and strategies was published in Van der Pool and Van Wijk 1995.

2. The overview Breetvelt presents can be summarized as follows (for references see Breetvelt 1991).
   The relation between the writing process as a whole and writing skills has been studied by, for instance, Stallard 1974, Perl 1979, Pianko 1979, Warters 1979, Atlas 1979, Rose 1980, Sweeder 1981, and Taylor 1984.
   Examples of studies in which one aspect of the process (planning, pausing, revising etc.) is studied in relation to writing skills are:
   - Flower and Hayes 1980 who focused on rhetorical problem definition;
   - Kennedy 1985, and Root 1985 who investigated idea generating;
   - Sommers 1980, Hayes 1981, Hayes 1985, Faigley 1981, Monahan 1982, Bridwell 1980, Garrett 1981, and Gee 1984 who focused on revision;
   - Schumacher 1979 who studied pausing;
   - Wright 1983 who investigated prewriting and revision;
   - Hull 1984 and Beach 1976 who studied editing/self evaluation.

3. The effect of such 'loosely'-defined or undefined concepts to weaken the explanatory power of a model is discussed for instance by Cooper and Holzman 1983, Cooper & Holzman 1985, and Stotsky 1990.

4. This may be due to the use of on-line data, the thinking-aloud protocols, as the process of planning and revision are observed at different moments. One could therefore assume that they are completely different.

5. This difference within the functioning of the conceptual processes relates to differences between *implicit* and *explicit* learning of skills (Ellis 1994). Implicit learning takes place naturally, simply and without conscious operations. Explicit learning is a conscious operation where the individual makes and tests hypotheses in search for structure.
   All kinds of non-linguistic sensory, motor, social and intellectual skills require a certain degree of control in order to function normally. This is acquired by getting experience by playing or 'modelling' etc. In principle, all language users have mastered these implicitly learned skills. In addition to these, there are explicitly learned skills that need just to be mastered in certain circumstances, and that can merely be acquired laboriously by specific study or training, when one has the capacities and is properly motivated. People learn to walk and to talk, for instance, but not all of them succeed in running a marathon, dancing a tango, presenting a paper, or singing an aria.

6. For a right completion of the figure, only the middle column of figure 3 should be included as the knowledge-telling process. The boxes referring to content and discourse knowledge should not be included. Otherwise it would result in an unjust duplication in figure 4.

7. Especially the writing models of the following researchers have inspired us in developing the blueprint: Bereiter and Scardamalia 1987, Flower and Hayes 1981, Cooper and Matsuhashi 1983, Witte and Cherry 1986, McCutchen and Perfetti 1982.

8. See also Scinto 1986 for differences between written and spoken language.

9. This study focuses on the production of text, not on generating isolated lexical or syntactic utterances such as words, idiom or single sentences. By text we mean a series of sentences not only coherent in relation to each other, but also in relation to the context (Halliday & Hasan 1976:23).

10. In some models it is assumed that the 'text-to-be-produced' is also mentally represented. In this model the text-to-be-produced is restricted to the writer's intention.

11. The role of intentions and expectations will not be discussed extensively. They are only mentioned as factors influencing the production process (see for instance Grosz & Sidner 1986).

12. The lexicon can be enlarged in the same way as world knowledge. For the sake of completeness it needs to be said that in figure 1 an arrow can also be drawn from the conceptual model to the lexicon. This arrow would symbolize the development of the lexicon, whereas the other arrows in the figure mean 'using' or 'sending' the information; the former arrow is not represented in figure 1.
In addition, it should be pointed out that the relations between the units depicted in the figure represent only the minimal ones, generalized over tasks. An additional relation might exist, for instance, between the *Invent*-component, the knowledge base and the lexicon; for instance a certain intention (e.g. 'at client's request inform about X') may activate a discourse scheme (e.g. a scheme for a business) as well as certain lexical expressions (e.g. a phrase like 'in response to your letter ...').

13. The perceptual loop does not necessarily take place after writing, nor does the whole loop need to be completed. It may also be used to change intermediary products such as the preverbal message (cf. covert repairs, Levelt 1989: 475-478). The perceptual loop presented in this model just indicates the way off-line revision takes place.

14. The way detection of errors and inappropriateness may take place is discussed in more detail by Bereiter and Scardamalia (1987: 266 et seq.). They assume that revision is executed according to a Compare-Diagnose-Operate principle, abbreviated as CDO.
The output of a process is checked on errors and inappropriateness by comparing it to the intended output and to general rules of language and communication (for instance spelling and syntax rules, Grice's maxims (Grice 1975), politeness rules (e.g. Brown & Levinson 1978). What the output is compared to exactly, depends on the level on which CDO takes place. This evaluation not only takes place at the end of the writing process. It also interrupts parts of the process (rereading, planning, generating etc.). In such a case, the CDO-process ends by returning to the interrupted process.
Bereiter and Scardamalia assume that the CDO-process is initiated by a perceived mismatch between two representations (1987: 266). However, they are not very explicit in what exactly is compared with what. Is every output compared to the intended text, and all knowledge? A second question related to the first is what format the 'intended text' has, and whether it always exists?
Absence of this 'text-to-be-produced', in the form of intentions, may explain for instance why many writers still have errors and inappropriate elements in their final texts. CDO cannot operate if there is nothing to compare. Another explanation for errors in final texts is that the CDO-process has not been executed or has been executed in a wrong way. Assuming that the operation disfunctions, it is unclear what the cause is: is it because not all output has been compared or is the disfunctioning caused by an inadequate way of comparing the output?
As the CDO-process still raises questions, it will be clear that this revising process needs to be explored in more detail.

15. This suggests that writers may find it considerably easier to detect problems in texts by others than in their own texts. This prediction was strikingly confirmed by two experiments by Bartlett (1982: 352).

16. The perceptual loop is related to monitoring language as it has a controlling function aimed at repairing errors and inappropriatenesses (Levelt 1989: 459). The role of monitoring in the writing process is limited in this study to the perceptual loop and it is only discussed in so far as it concerns the conceptual component of *Inventing*. For monitoring in-between products, see also Levelt (1989: 470, figure 12.3).

17. We refer to *Reflect*, *Select*, and *Shape* as 'operations' or 'functions' and not as 'processes', as we make no claims about their input, output and status.

18. For (recent) research on planning of text see for instance Günther 1993; for a critical view of the concept of 'planning' see Stotsky 1990; for research on a particular part of planning, that is pre-text, see Witte 1987.

19. Levelt distinguishes between covert repairs and overt ones. In covert repairs it is unknown what has been repaired. The overt repairs include a traceable, that is an already uttered, troublesome item (1989: 478-479). This distinction also holds for writing. The overt repairs can be observed in crossed out words, the covert repairs are hidden in the writer's mind as they occur before motor activities have taken place.

20. As we prefer labelling in terms of the type of process, for reasons presented in section 2.2, the terms *Guide* and *Control* are used instead of 'planning' and 'revision'.

21. The control function of *Reflect* is not included in the example.

22. Following Levelt (1989: 110), a distinction is made between communicative intentions and illocutionary intentions. The latter are a subdomain of the former.

23. When the linguistic processes are automatized, processing resources become available for conceptual processes (Van Dijk & Kintsch 1983: 24; Bossers 1993a,b). Only if processes take place with great difficulty, for instance by a poor control of language or a physical disorder, a language user may have recourse to temporary solutions in the form of compensation strategies, such as avoiding strategies (see Kolk's adaptation theory of agrammatism 1987).

24. Frequently used synonyms for *strategy* are 'heuristic' and 'management'; synonyms for *procedure* are 'rule' and 'implementation'. The combination of goal and strategy is frequently indicated as 'plan'; the combination of strategy and procedure is often indicated as 'method'.

25. See for instance: Hull 1984, Sommers 1978, Bereiter and Scardamalia 1982, Flower and Hayes 1980. For a comparison between 'remedial' and 'mean' writers, see Pianko 1979.

26. This dichotomy does not mean that all experts' behaviour can be unconditionally characterized in this way. Experts merely have the *ability* to do so. Their writing experience has delivered a wide repertory of strategies leading to clear procedures. They are therefore capable of weighing the strategies and procedures and of making their choice on the basis of their intention(s).

27. It is not surprising that novices mostly use less knowledge in writing, as they probably have less knowledge because of their young age. This may be ascribed to general cognitive development.

28. This characterization (writer-based prose) is dominated by the way the knowledge is represented in the writer's memory (Bereiter & Scardamalia 1987: 184). Britton et al. use the term 'expressive writing' as a kind of writing directed by internal needs (in contrast to 'transactional writing').

29. Bereiter and Scardamalia (1987) found that only if children were supported to evaluate every segment immediately after production, did they appear to be able to revise their texts ('Alternating Procedure', 1987: 282).

30. Bereiter and Scardamalia do not assign age norms. Their main reason for this is that they are interested in the nature of conceptual processes and how they develop (1987: 211-212).

31. In a school context such a problem is narrowed by an assignment, which is a restriction and not a complete specification of an intention.

32. In this study the focus is on the individual writer rather than on aggregated data in which individuals are grouped together from the start on the basis of predetermined categories.

3 Methods to study conceptual processes

3.1 Introduction

The study of conceptual processes in writing runs up against at least one major methodological difficulty (Faigley et al. 1985: 163). The conceptual processes are very difficult to get access to directly, as they are hidden in the mind, which restricts the possibilities for experimental manipulation. Nevertheless, there is an indirect way to gain insight into these processes. The hidden processes lead to observable manifestations during and after writing, such as pausing, noting, nibbling at one's pencil, and of course, the text itself. Analysis of these manifestations provides a key to the covert processes. A constraint, however, is that the manifestations should be of one and the same kind. In order to control variation between manifestations, the task should be well-considered (for task effect, see e.g. Van den Bergh & Meuffels 1993).

This chapter deals with the methodological issues related to cognitive writing research in general, and to this study in particular (see the marked boxes in figure 1). It motivates the choice of textual manifestations, and the concrete form of these off-line observations, that is to say the corpus used in this study.

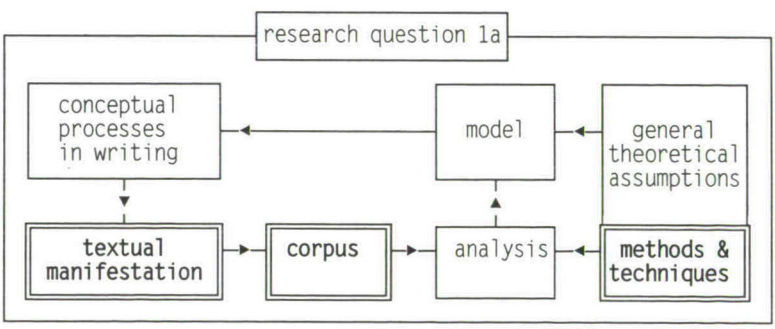


Figure 1. The position of this chapter in the study

A short overview and characterization of research methods often used in writing research is presented in section 3.2. As this study is a text-analytical one, special attention will be paid to the characteristics of text analysis (section 3.3), in particular the relation between analyzing

and reading - as these have much in common -, and to the way the results of text analysis can be used to infer process aspects of writing.

The writing assignment that yielded the corpus of texts investigated in this study is described in section 3.4. The original corpus and the sample analyzed in this study are characterized in section 3.5.

3.2 Methods

The observable manifestations of conceptual processes can be categorized with respect to the moment at which they are obtained and the extent to which they are intrinsic to the writing process. These manifestations can be obtained during the process itself (*on-line*) or afterwards (*off-line*). They may be a natural consequence of the process (*intrinsic*), such as pauses or texts, or they may be explicitly elicited (*extrinsic*), for instance, by asking subjects to express their own thoughts.[1] A schematic representation of frequently used types of manifestations is presented in table 1.

Table 1. Types of manifestations of conceptual processes and an example per type[2]

	extrinsic	intrinsic
on-line	thinking aloud	pauses
off-line	interview	text

In regard to research on conceptual processes Ericsson and Simon (1993:7) state that "subjects have the sole direct access to their own mental states and processes". That is probably why many studies have heavily relied on asking subjects to verbalize their thoughts during writing (e.g. Emig 1971, Flower & Hayes 1980). In the **thinking aloud** method a subject is asked to verbalize everything that is going on in his mind during composing. These uttered thoughts are recorded on tape and written out in protocols for further analysis (see chapters 6 and 7 in Ericsson & Simon 1993).

Although the method is intrusive, and forces the writer to speak and to write at the same time, the protocols are reliable means to provide evidence about the course of the ongoing task. According to Ericsson and Simon (1993: 220) language producers are not "schizophrenic creatures who produce a stream of words, parallel but irrelevant to the cognitive task they are performing". The thinking aloud protocols shed a special light on the execution of a task,

because of this relation between linguistic encoding and the conceptual activities. Information which is missing in other research methods may be (partly) verbalized in thinking aloud; by revealing this information the writers can provide a picture of strategies employed and inferences drawn (Ericsson & Simon 1993: 220).

It is likely that the informativeness of the protocols depends to a high degree on the developmental level of the subject. The technique requires a lot of meta-cognitive and formulation skills (Cooper & Holzman 1983; Smagorinsky 1989; Lowijck & Vanmaele 1992). If these skills are not (yet) sufficiently developed, subjects have great difficulty executing the task, for instance because they have insufficient formulation skills and reflection on the procedures that are going on. As Bereiter and Scardamalia state, children may not be able to reveal as much of their thought while composing as adults may do (1987: 70).

In addition, thinking-aloud protocols need to be analyzed. There are several programmes for encoding and analyzing these protocols, but more developmental effort will be required (Simon & Ericsson 1993: 305). Many researchers are wrestling with problems of categorizing utterances (e.g. Cooper & Holzman 1983: Faigley et al. 1985; Van der Geest 1991: 181 ff.). In order to increase the reliability, two or more coders in general categorize the utterances (Andringa 1990: 253). Nevertheless, such an analysis still presents difficulties. Bereiter and Scardamalia (1987: 206) indicate that judging protocol statements largely depends upon subjective criteria. It is often uncertain in judging protocols whether a statement is meant to describe the composition process (meta-textual), or is meant to be included in the text (textual). The utterance "The main point I want to make" (Bereiter & Scardamalia 1987: 203) is a good example illustrating this diffuse distinction. This utterance can be either meta-textual or textual. In the first case it will be regarded as a description of the process, and classified as an instance of conceptual planning related to a *Reflect* activity. In the second case it is regarded as being included in the text, and classified as a manifestation of content generation, related to *Select* activity.

Another extrinsic method is **interviewing**. In this off-line method subjects are asked (or have) to report immediately after writing about salient aspects of the process.[3] The advantage of technique is that the meta-cognitive task does not interfere with the writing process. Interviewing does not influence the cognitive load during writing, in contrast to thinking aloud. Still, the method relies on the subject as a direct source of obtaining data, and in reporting about cognition on a meta-level a subject needs to master meta-skills. This may be one of the reasons why this technique has often been applied in research that has literary writers as subjects.

Registration of pauses does not depend on these meta-cognitive skills. Instead of asking subjects to reflect on their own thought processes (during or after writing), the subject's on-line writing behaviour is observed. The observable writing process consists of intervals of motor activities alternated with pausing. This method detects transitions between writing and not writing, i.e. thinking.

The claim underlying this method is that the motor activities correspond to *Execute* processes, whereas the pauses mirror *Invent* processes, in specific planning (Matsuhashi 1981). It is assumed that during pauses subjects reflect on what they have written and/or what they will be writing. The length of the pauses provides information about the *duration* and probably about the *extent* of the reflecting (Good & Butterworth 1980; Matsuhashi 1981). The production pauses are therefore observable indications of where and how long a subject pauses in order to plan.

These registrations can be done very precisely with the present technical facilities. The obtained data, however, provide direct insight into the activities of writing only on a surface level. They do not provide unambiguous insight into processes and thoughts.

Although pause registration does not depend on the subject's meta-linguistic skills, it has another problematic side. It is based on the claim that what is done during pausing represents (parts of) the writing process. And that is exactly what the problem is with young writers. It is hard to assume that novices, if they are not writing, are engaged in mental activities strictly related to producing text. Especially these young writers may be distracted by events or objects around them, such as snow falling, or other children playing outside etc. The data might be corrected for these distracters, and of course the noise is (probably) present in all conditions. Nevertheless, it will still be hard to maintain the claim that what is observed during pausing by novices is exclusively related to text production. As a pause may indicate a heterogeneous set of processes, pause registration is not very reliable for studying writing processes of novices.

In addition to pauses, the writing process yields another off-line observable manifestation, which is text. As texts include clues to the conceptual processes (Bereiter & Scardamalia 1987: 41), **text analysis** should provide a suitable basis for revealing cognitive aspects (Cooper 1983; Frederiksen 1986; Matsuhashi & Quinn 1984; Tamor & Bond 1983). Whereas pauses provide information about the *duration* and the *extent* of the processes, texts provide insight into the *characteristics* of underlying cognitive processes.[4]

As the aim of this study is to gain insight into the development of conceptual processes of text production, the writing processes of experts as well as those of novices have to be studied. This, obviously, restricts the choice of research methods. For a reliable comparison between

observations of novices and advanced writers, meta-cognitive skills ought to be avoided, and it should be clear that the observations concern cognitive *writing* processes. Therefore it was decided that text analysis should be used as a research method in this study. It sidesteps metacognitive skills, and allows inferring the characteristics of the conceptual processes on the basis of observed features (see also section 3.3.2).

In order to indicate the relevance of text analysis with respect to other methods in writing research, the levels of inquiry distinguished by Bereiter and Scardamalia are illustrative (1987: 35 et seq.). Bereiter and Scardamalia categorized research on writing processes into six interacting and cumulative levels.[5] The levels are not merely sequential but arranged in a dimension of abstractness. Bereiter and Scardamalia remark that "understanding of the composing process does not emerge from inquiry on any particular level but rather synthesizing knowledge gained in the course of spiralling through levels" (1987: 33).

The basic level comprises Reflective inquiry. It forms the home base, as it is the place from which other kinds of inquiry start. On this level the central questions are 'what is the nature of the current phenomenon?', 'what are the problems?', and 'what do the data mean?'. This is where research keeps returning for fresh starts, where knowledge finally becomes understanding.

This reflective inquiry is enriched (not replaced) by what Bereiter and Scardamalia call 'empirical variable testing' on the second level. It borrows its name from the fact that testing a premise empirically requires translating the premise into a statement about the relationship between observable variables. In the case of a premise such as 'a pupil who reads good literature writes well', one has to operationalize the concepts 'good literature' (for instance as the books listed by a prominent literary critic) and 'writing well' (for instance as ratings by a teacher). The premise has to be translated into observable entities or behaviour. These dependent variables are subsequently tested in order to refute or confirm the premise.

This empirical testing opens the door for the third level of inquiry, which is *text analysis*. Texts are the input for this kind of research. The aim of analyzing texts is to describe the texts in terms of rules or principles, and not just empirically in terms of variables that may be derived from the texts, such as a number of subordinated clauses or the number of connectors.[6] "Level 3 inquiry approaches texts as complex phenomena that exhibit internal lawfulness, and it aims to understand that lawfulness" (1987: 40)[7]. Such a characterization of texts in terms of rules and principles does not represent the composing process yet, but it is vital in that it can provide indirect evidence. It answers questions like 'what makes this text seem what it seems to be?', and 'what rules does the writer obey?'.

In order to gain insight into the patterns of the writing process, it is necessary to extend text analysis with a description of the process. Questions like 'what is the writer thinking?' and 'what pattern or system is revealed in the writer's thoughts while he is composing?' should be answered on this fourth level. This 'process description inquiry' searches for lawfulness in the conceptual processes, and results in a theory or model. Of course, the methodological problems are great on this level because the processes are not directly accessible. Therefore Bereiter and Scardamalia give a warning not to confuse level-4 inquiry with the methods that are normally used to conduct it (1987: 42). "The defining characteristic of level-4 inquiry is its search for a description of the composing process. Because most of the process is covert, Level-4 researchers tend to get involved in mind-reading problems of some sort, but this is a secondary problem. Level 4 researchers could limit their study to the observable part of the composing behaviour, obtaining a valid, although necessarily limited, description of the process (e.g. Stallard 1974).".[8]

On the next level, it is tested in a highly controlled way whether a theory or model holds for specific circumstances. The theoretical construction which is the result of preceding levels is tested for its empirical implications. This so-called 'theory-embedded experimentation' is used for testing assumptions of theories or differences between theories.

In order to represent the (parts of) the composing process or to test how well theories and models represent it, extension with level 6 is necessary: 'simulation' by the computer. This inquiry concerns the investigation of the nature of different composing strategies or abilities by trying to simulate them.[9]

This study is not restricted to text analysis in the strict sense of level 3 inquiry. The patterns that result from the text analysis are related to conceptual processes as depicted in the model presented in chapter 2 (section 2.3). By modelling these results, an attempt is made to gain insight into the characteristics and development of *Reflect* and *Select*. With this approach we aim at a modest contribution to a *child rhetoric*, which means "working out a rule system that gives rise to texts like those children write" (Bereiter & Scardamalia 1983: 23).

3.3 Text analysis

Text analysis covers a multitude of methods for analyzing linguistic expressions. Dependent on the unit of analysis, it may vary from analysis on clause level to text level. As the aim of this study is to investigate cognitive processes of *text* production, text analysis should be taken literally in this case, meaning analyzing complete texts.

This kind of text analysis is closely related to reading. The same kinds of operations take place, albeit more explicitly and with a different aim. Because of the similarities, it is informative to have a closer look at how reading and analysis are related. This excursion is presented in section 3.3.1.

Text analysis unravels linguistic features. Textual aspects that are most informative and can reliably be interpreted and used to model conceptual processes are discussed in section 3.3.2.

3.3.1 Relation to reading

The main difference between analyzing and reading a text is the fact that analyzing is a meta-linguistic activity whereas reading is a (communicative) linguistic one. An analyst consciously unravels a text in order to make (text) features explicit.[10] An analyst and reader, however, make use of equivalent processes. Therefore it is informative to take a closer look at the reading process.

A schematic representation of the reading process is depicted in figure 2. It is based on general theories of reading (Dijkstra & Kempen 1993; Balota, Flores d'Arcais and Rayner 1990), and more specific models and theories about written text processing (among others Kintsch & Van Dijk 1983; Rayner & Pollatsek 1989; Just & Carpenter 1987). The components correspond to the ones depicted in the right part of the blueprint for a writer (see chapter 2, figure 5; Van der Pool & Van Wijk 1995).

The components of the reading process are depicted from the bottom to the top of the figure 2. The process starts with *perceiving* a series of characters from the text by means of sensory activities of the eyes, which results in a *graphemic string*. *Decode* processes reconstruct the lexical content and the syntactic form of the clause-like elements from this graphemic string, which yields a syntactically and semantically *parsed string*. Such parsed strings are connected to each other and enriched with world knowledge in the *comprehend* component. This conceptual operation ends in a mental representation stored in the text record (see for instance Noordman & Maes 1993).

As the different operations within the conceptual component play an important part in understanding texts, and therefore also in analyzing texts, this component is discussed in detail.

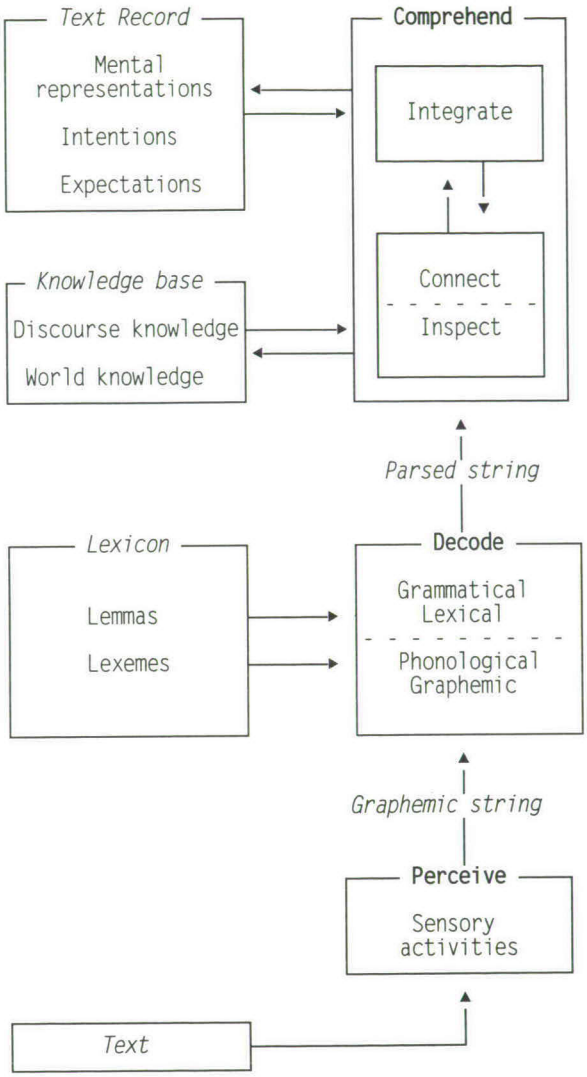


Figure 2. Organization of reading processes

Comprehend

The operations within the *Comprehend* component correspond to Kintsch and Van Dijk's model of comprehending text (1978; and Kintsch 1988). When a sentence of a text enters this component, the underlying propositions are analyzed. Subsequently, it is determined how these propositions can be connected to a previous one.

This propositional analysis holds for every incoming sentence of a text. If it is not possible to link a sentence directly to a previous one, a more remote proposition is sought or a new one is constructed. In addition, on a higher level sets of propositions are connected to each other in order to create a coherent mental representation of the text.

Comprehend consists of three operations, each corresponding to the activities mentioned above. We label them as INSPECT, CONNECT, and INTEGRATE. In understanding a text, a reader searches within a clause for indications, implicit or explicit (INSPECT) in order to link that clause to other clauses (CONNECT). These interclausal relations may be established among other things on the basis of referential expressions (pronouns, nominal anaphors), or relation markers such as connectors. On the basis of these linguistic markers CONNECT constitutes local coherence.

Comprehending a text, however, is more than being aware of the direct interclausal relations. A text is not just a sequence of interrelated clauses; it is coherent on a global level as well. This global coherence is established in INTEGRATE. It enriches the connected clauses with world and discourse knowledge to create a coherent mental representation of the text on all levels, including the global one. Stereotypical text structures, such as frames, scripts and superstructures, direct the linkings of the clauses on a local as well as on a global level (see for instance Schank & Abelson 1977).

In addition to this guiding function, INTEGRATE also checks the local connections on the basis of the global coherence. If an ambiguous anaphor, for instance, leaves open two possible linkings in CONNECT, INTEGRATE decides on the basis of world and discourse knowledge which one will fit best.

The comprehending process is guided by reading intentions and expectations. The intentions determine to a great extent which information is remembered. If one reads a description of a house with the intention of buying it, for instance, one will remember different information than if one intends to burgle (Anderson & Pitchert 1987).

The expectations are mainly based on knowledge and acquired experience in the form of stereotypical patterns such as scripts (Schank & Abelson 1977). They are, among other things, activated by specific phrases or expressions. After a phrase like 'Once upon a time', for example, a reader expects that a fairy tale will follow. Subsequently, such a phrase activates expectations about the content and structure of the story: the main characters are introduced (persons, animals or imaginary creatures), 'good' and 'evil' clash, but fortunately everything turns out right, and so they live 'happily every after' (Propp 1968).

To illustrate the operations in the Comprehend component, imagine someone reading a fragment (1) in an arbitrary book in a bookshop.

(1) a. The crew entered the platform
b. and installed themselves in the narrow cabin.
c. The engines made a deafening noise at the start.
d. Within one minute they were out of sight
e. and silence returned.

How does he comprehend what is described? A possible characterization of the conceptual operations is outlined in table 2, which is an example of the type of operations involved in understanding a text.

It should, however, be remarked that this example does not indicate anything about the temporal course of these operations. For the sake of completeness, it must therefore be added that the comprehend operations are cyclic, in the sense that one runs through them for every new input. Consequently, it means that the three operations interact top-down as well as bottom-up. INTEGRATE guides and controls INSPECT and CONNECT in a 'top-down' way, whereas INSPECT and CONNECT influence INTEGRATE in a 'bottom-up' way.

It would be beyond the scope of this study to discuss this temporal course and related interaction of the operations in more detail. The example is only a limited characterization of the type of operations involved.

In order to discover coherence between the individual segments of text fragment (1), the segments are **inspected** for the presence of linguistic features, as they can provide hints to **connect** the segments. The contraction in (1b), for instance, indicates a relation with (1a); the verb type (action verb) and the past tense indicate that it is probably a series of events which is described.

On the basis of such features, a reader may have determined coherence between *separate segments* (intention IIa), but that does not mean that he knows what the *text* is about (intention Ia). A text is, after all, more than a list of connected segments (Hobbs 1990). In order to understand the fragment, intentions IIb and IIc need to be achieved as well. Thus, the reader has to look for shared features in the clauses, and relate the connected elements to his world and discourse knowledge. A feature that the clauses in the fragment share is that each clause describes an event; this series of events may indicate a narrative genre. But with respect to expectation IIIb there is more. The sequence of events was expected to be about space travel. A reader concludes, on the basis of world knowledge and the bottom-up input of INSPECT, that

words like 'crew', 'platform', 'cabin', 'start' and 'deafening noise of engines' suggest the launching of a missile rather than car racing.

Table 2. Conceptual operations in reading a text fragment of an arbitrary book

Integrate:	*What are my communicative intentions?*		
	Consider possible communicative intentions		
	Ia	I want to know what the text is about	
	Ib	I am searching an instruction about motor techniques	
	How can I achieve that with this text?		
	Specify intention:	IIa	determine coherence between subsequent segments
		IIb	look for shared features (topic, genre)
		IIc	relate text, if possible, to world knowledge
	What are my expectations?		
	Consider possible expectations		
	IIIa	It will be about car racing	
	IIIb	It will probably be about space travel	
Connect:	*How can I connect segments?*		
	Consider possible arguments for connecting		
	IVa	presence of a conjunctor	
	IVb	semantic type of main verb	
	IVc	problem/goal discussed	
	Determine connection (on the basis of intentions and expectations):		
	Choose ..... (this differs per segment)		
Inspect:	*Which linguistic elements serve as indicator for connecting?*		
	Consider possible indicators for connecting segments		
	Va	sentence structure: contraction, position clause...	
	Vb	references: type and position of anaphors....	
	Vc	word semantics: action verb, sequence-builder...	
	Determine indicators (on the basis of, among other things, text constraints, reading goals):		
	Choose ..... (this differs per segment)		

By **integrating** segments into world knowledge, they are put into a meaningful perspective (achievement of intention Ia). It becomes clear, for example, that 'they' in (1d) does not refer to the closest referent 'engines' in (1c), or to the singular word 'crew' in (1a), but to the unmentioned missile including the astronauts. The ambiguity of connectors may be solved by integration with world knowledge as well (Sanders, Spooren & Noordman 1993). For instance 'and' in (1e) is not only sequential ('and then'), but also causal ('and that is why').

3.3.2 Inferring process from product

Although analyzing consists of operations that also apply in reading, it has a different aim. Analyzing a text is mostly a means to create a meta-linguistic description of (parts of) the text, in a more explicit and systematic way than in reading. It is an instrument to put (parts of) texts into comparable formats, in order to explore and describe differences between the texts.

That purpose requires the method to be reliable. It should prescribe what features on what level have to be *inspected* and what relevance they have with respect to the *connection* of segments. In addition, for a coherent *integration*, the extent to which an appeal is made to world knowledge is crucial. A method of analysis should therefore be *explicit* and *procedural*, in order to minimize the analyst's intuitions.

The method of analysis should be based on explicit definitions of the linguistic features it inspects, and procedural rules for the analysis. It has to be clear which features are relevant, and which are not. The same holds for rules that take care of connecting segments. And last but certainly not least, the appeal to world and discourse knowledge has to be specified as much as possible.

Most methods are quite explicit in defining the features that are analyzed, but less explicit in the appeal to world knowledge. The use of knowledge is very important to give an adequate description, especially in analyzing on a discourse level. This is mostly referred to as 'the analyst's intuitions'. Mann and Thompson (1988: 245-246) remark that an "analyst has access to the text, has knowledge of the context in which it was written (...), but has no direct access to either the writer or other readers. During analysis, judgements must be made about the writer or readers. Since such judgements cannot be certain, they must be *plausibility judgements*.[11] In effect, every judgement of the completed analysis is in the form, 'It is plausible to the analyst that ...'". Especially an unverifiable appeal to intuitions decreases the reliability of a method.

Text analysis as a means to provide insight into conceptual processes is based on two constraints. First of all, - as was stated above - the instrument should yield a reliable result which reduces texts to comparable units. Secondly, the correspondence between the analytical results and the conceptual processes should be established before processes can be inferred from the products. This raises the question what clues in a text reveal (parts of) the writing process. An integration of prominent psycholinguistic ideas about this correspondence can be summarized as follows.

The main components of the writing process, as described in chapter 2, are *inventing* on a conceptual level, *encoding* on a linguistic level and *executing* on a physical level.

If we assume that these operations leave their traces in a text, the content and structure of a text mirror what information has been retrieved and organized, and reflected on (Scardamalia & Bereiter 1987: 206). Similarly, it can be assumed that the information distribution results from shaping (Levelt 1989: 99). However, there is less evidence that sentence structure and choice of words have an unambiguous one-to-one relationship with linguistic encoding. The preverbal message, the output of *Invent*, also influences sentence structure and choice of words. Inventing a message probably triggers words and formulations (Levelt 1989: 200). The precise way in which this takes place is unknown (see Levelt 1989: 163, 179). The same holds for the Execute component and its correspondence to text features. The relations between text features and conceptual processes are depicted in figure 3.

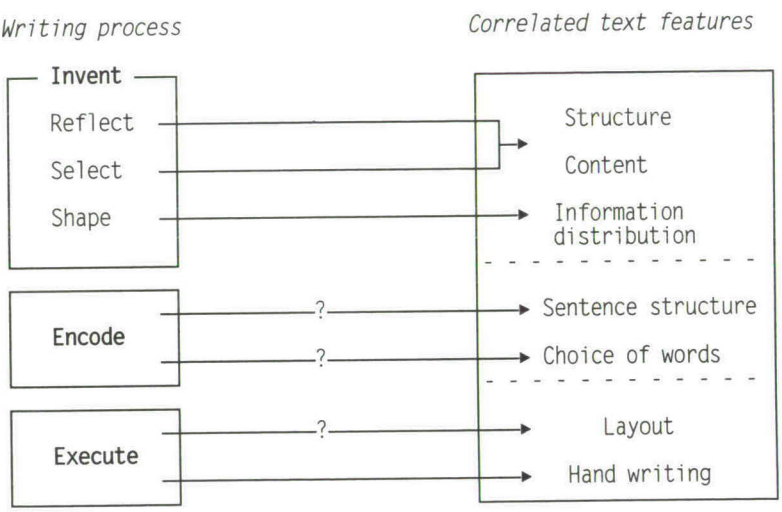


Figure 3. The relations between process components and text features

The above correspondence between conceptual operations and textual features shows that features on a text level seem to be the best indicators for conceptual processes. As this study focuses on writing as a conceptual process, it is necessary to take a closer look at the relationships between the operations of *inventing* and its corresponding text features. This relationship is depicted in table 3.

Table 3. Conceptual operations related to text features

Invent	Text features
Reflect	
- Guide Select and Shape: Consider, determine and specify intentions and expectations (e.g. convincing, amusing)	- type of information (e.g. descriptive, narrative) - content and global text structure
- Control Select and Shape: check agreement with intentions and expectations; if necessary adjust to intentions and expectations	- comments and revisions based on intentions and expectations
Select	
- retrieving information	- propositional content
- organizing information	- local text structure
Shape	
- propositional shaping	- given-new distribution, focus, topic

In Reflect the communicative intentions and expectations are examined, determined and specified. They are used to guide and control Select and Shape operations (see also chapter 2). As regards the communicative intentions, the prime question is what one wants to achieve with the text. It is important to make these intentions explicit, as they function as guidelines for the retrieval and organization of information. The intention of convincing a reader, obviously, requires a different selection of information than amusing, for instance. In addition, the intention will influence the organization of information (e.g. placing the best arguments in the most prominent position). Thus, the guiding functions of Reflect, although more or less internalized, will most clearly leave their traces in the type of information, the content and global structure of the text. The control operations of Reflect, may be covert as well; they are therefore hardly noticeable in the final text (cf. covert and overt repair, Levelt 1989: 478-479). First drafts, however, may include overt repairs in the form of comments and revisions that are explicitly made in order to adjust (or 'repair') previously given information.

The operations that most unambiguously leave their traces in a text are the retrieval and organization of information. The result of retrieving is the propositional content. Organizing information manifests itself in the global and local structure of a text. So, content and text structure mirror the Select-operation (Scardamalia & Bereiter 1987: 206).[12]

Once it has been established what information will be included, the next step is to put the information into a certain shape. One has to determine how one wants to present the information, i.e. considering the distribution of given and new information, the elements to be put in initial position of the sentence, the type of references to use etc. These textual features therefore bear traces of the shaping operation.

In studying Reflect and Select processes, as is the aim in this study, content and (text) structure may be most informative about these Invent operations. Text samples, however, mostly display an unlimited variation in content, and the way this content has been organized and formulated. When different persons write a text, the resulting texts will be different, even if topics, instructions and skills are more or less the same. Two texts are never completely alike, therefore they can only be used as data to infer processes if they can be compared with other texts.

In contrast to the analysis of the content, a structure analysis is more or less independent of the topic and the exact wording of the text. This is especially important when imperfect texts written by inexperienced writers are analyzed. Analysts examining the same text should come to more or less the same conclusion. As the content of a text is more sensitive to different interpretations than the structure, analysis of text structure provides more reliable results. In this study the analysis of text structure is therefore used to infer conceptual operations from products. The analytical method used in this study will be described in detail in chapter 4.

3.4 The writing task in this study

Research on developmental aspects of the conceptual processes concerns analysis of texts written by novices as well as by more advanced writers. Writers varying in age and skills must produce a text within their own (cognitive) possibilities, which obviously restricts the possible tasks. The novice writers determine the upper limit, as they must be able to perform the task as well. In order to meet this requirement, the writing process, and especially the conceptual part, has to be fairly uncomplicated. Thus, the choice of the text genre to be produced plays a crucial part in determining the writing assignment.

The most uncomplicated way to produce a text is to *describe* a phenomenon. In describing a topic, one is faced with the "problem of content selection" and the "problem of content organization" (Mosenthal 1985: 392). These problems are of a different nature than those encountered when writing an instructive or an argumentative text. The problem of retrieving

in describing is reduced to observation of already stored information. Writing a descriptive text does not take (much), if any, argumentative reasoning, as organization is easily achieved (Applebee 1978). Thus, the most uncomplicated way to invent a message is to write a descriptive text (Christensen 1967, Virtanen 1992).[13] From a developmental point of view Christensen (1967: xi) argues that the best form of writing is description, because it is close to the child's experience; it is concrete and the problem of invention reduces itself to observation.

In order to be able to study such relatively uncomplicated writing processes, the decision was made to use descriptive texts. The assignment used in this study is presented in table 4.[14]

Table 4. Specification of writing task

Assignment:[15]

> Describe the person you would like to resemble. It may be a really existing person as well as an imaginary one; or someone that incorporates the characteristics you have noticed in different persons.
>
> Tell something about that person: How old is he? What character does he have? What does he look like? What is his profession? Where does he find recreation, and what does he do in his spare time?
>
> If you write about an existing person, please write down "existing person" at the end. You do not have to mention his name if you do not want to.

Constraints of the assignment:

1. Describe a person and discuss the five themes specified (*Select*)
2. Evaluate the choice of this specific person (*Reflect*)

Subjects were asked, during normal school hours, to describe the person they would like to resemble. This ideal person could be an existing or an imaginary one. They were asked to tell something about his/her age, character, physical appearance, profession, and hobbies. Texts were written spontaneously with only little editing (crossing out of words).

The assignment had two constraints. The first, explicitly stated in the assignment, concerns Select-operations. The writers had to describe characteristics of the person they would have liked to resemble and discuss the five relevant themes which were given in the instruction (in

the form of questions). These were, in the order of appearance in the assignment: Personal particulars (age), Character, Appearance, Career and Hobby. These themes should serve as cues for the writers in retrieving information. The question is how writers retrieved and locally organized the information.

The second constraint, which could be inferred from the instruction, relates to Reflect-operations: describing a desire like 'would like to be like' invites for an evaluation of the choice of that particular person. The question is whether writers interpret the task in this way as well.

This writing assignment was presented to pupils of primary and secondary schools in Europe as part of a large-scale European psychological research project on the development of self-concept executed in 1964 and 1965 (Lutte, Mönks, Kempen & Sarti 1969). In the Netherlands and in Belgium 3450 pupils were involved in this project. They originated from 3 regions, Southern Limburg, Eastern Brabant, and Flanders, from each of which 1150 pupils participated. The corpus used in this study is a sample from the originally collected set of essays.[16]

In order to get a broader view on development, this assignment was presented to 12 adults as well.[17] The subjects were advanced university students of the Department of Language and Literature in Tilburg.

3.5 The corpus

This section presents a preliminary exploration and description of the original corpus in order to be able to take a representative sample (section 3.5.1). Subsequently, the sample central in this study will be described on the basis of descriptive syntactic indices. These results will be compared with other studies on this topic to check the representativity of the sample (section 3.5.2).

3.5.1 Preliminary exploration

A sample of the corpus was used before in a psycholinguistic study. Van Wijk and Kempen (1982a,b) studied a sample of 280 texts. The study focused on developmental patterns in syntactic features. Subjects came from the southern part of the Netherlands (the province of North-Brabant) and the northern part of Belgium (Flanders). They used a sample stratified on type of school, age and sex.

This research was successfully finished; several systematic changes in the syntactic complexity of the sentence structure were found (Van Wijk & Kempen 1982a,b). Further research, however, cast some doubt on the purely syntactic explanation of the developmental changes. The observed changes in sentences could have been caused by underlying conceptual factors (Van Wijk, 1987: 28 ff.). This idea was supported by results of a supplementary experimental study (Van Wijk & Luiten 1987), which showed that in selecting samples not only *subject features*, such as sex, age and type of school, should be controlled but also *text features*, such as topic and task interpretation. These text features had to be taken as a independent variable as well, for they might influence the conceptual processes. In the current study these conclusions have been taken into account. Thus, before a sample was taken a preliminary exploration took place.

From the corpus of 3450 texts written in Dutch, only those texts written by pupils at the ages of 10, 12, 15 and 16 (N=539) living in the region of Eastern Brabant and Flanders were explored. The subjects, boys and girls, were pupils of primary school grade 4 and 6, and of secondary schools (LBO, ULO, and Grammar school[18]) grade 3. The reason for selecting these pupils is twofold. First, their use of language is in conformity with the standard language, in contrast to pupils living in Southern Limburg. Second, this group was highly homogeneous as far as their socio-economic backgrounds are concerned.

The reason for studying the age groups 10, 12, 15/16 in particular is motivated by the developmental perspective of this study. In general, 10-year-old children are able to write a complete text; before that age they are hardly able to. The age of 12 marks the end of the primary school period. This last grade of the primary school consists of pupils who differ greatly in knowledge and skill. Future LBO-pupils as well as future Grammar School pupils are all in the same grade. The age of 15/16 marks the end of compulsory education. Until that age every pupil has had an education, after that age there is strong variation.

The texts of these 539 subjects were controlled for:
a. *subject* features, i.e. sex, age, and type of school;
b. *text* features, i.e. realization of the assignment, and interpretation of the assignment.

a) Subject features
The texts (N=539) are classified on the basis of the subject features of age, type of school and sex. This classification is presented in table 5.

Table 5. The total number of texts in relation to age, type of school, and sex

age	type of school	girls	boys	total
10	primary school	49	35	84
12	primary school	94	60	154
15	secondary (LBO)	56	49	105
15	secondary (ULO)	47	45	92
15/16	secondary (Grammar school)	44	60	104
Total		290	249	539

b) Text features

The texts were also controlled for interpretation of the assignment. The subjects were asked to describe a person they would like to resemble. A number of subjects (N=88) did not comply with this assignment. Some of them did not produce any text at all, or wrote a text shorter than three lines. Others gave an enumeration of characteristics without formulating complete clauses, or wrote a narrative without describing a person. There were also subjects who wrote about themselves, and other subjects who described several persons instead of one. These texts were excluded from the sample and are referred to as 'drop-outs'. The number of excluded texts per group is presented in table 6 in absolute numbers and in proportions.

Table 6. Number of 'drop-out' texts in relation to age, type of school and sex (proportion between brackets)

age	type of school	girls	boys	total
10	primary	6 (.12)	2 (.06)	8 (.10)
12	primary	16 (.17)	8 (.13)	24 (.16)
15	secondary (LBO)	7 (.13)	6 (.12)	13 (.12)
15	secondary (ULO)	9 (.19)	7 (.15)	16 (.17)
15/16	secondary (Grammar school)	9 (.20)	18 (.30)	27 (.26)
Total		47 (.16)	41 (.16)	88 (.16)

The proportion of drop-outs did not differ for the sexes (both 16%). The proportion of drop-outs per group is statistically equal for all groups.[19] All these drop-out texts (N=88) were excluded from further exploration.

Although the remaining 451 subjects complied with the assignment, a closer examination showed that not all of them had interpreted the assignment in the same way. The following five interpretations could be distinguished:

a) describe an existing person;
b) describe an existing person and attribute ideal aspects to him;
c) describe a constructed person and illustrate the description with one or more characteristics of an existing person;
d) describe a constructed person on the basis of characteristics of different existing persons[20];
e) describe a constructed person without reference to existing persons.

Table 7 presents the number of different interpretations of the assignment (a to e) per group in absolute numbers and proportions. Because the interpretations b, c, and d occurred infrequently, they were lumped together.

Table 7. Interpretations of the assignment in relation to age, type of school and sex

		girls				boys			
age	type of school	a	bcd	e	total	a	bcd	e	total
10	primary	.91	.05	.05	43	.91	.03	.06	33
12	primary	.78	.08	.14	78	.98	.00	.02	52
15	secondary (LBO)	.79	.10	.10	49	.84	.05	.12	43
15	secondary (ULO)	.74	.10	.16	38	.89	.05	.05	38
15/16	secondary (Grammar school)	.40	.20	.40	35	.40	.12	.48	42
Total		.74	.10	.16	243	.79	.05	.16	208

The majority of the subjects interpreted the assignment according to the a-version 'describe a concrete person'. However, this preference decreased with age from over 90% to about 40% (girls: $\chi^2(4)=29.05$, $p<.001$; boys: $\chi^2(4)=58.75$, $p<.001$). The major jump is found between 15-year-old secondary (ULO) pupils and those of the Grammar School.[21]

Closer examination of the texts in the a-category showed that two kinds of concrete persons were described:

- acquaintances from the private neighbourhood of the subjects, for example a mother, a brother, a schoolfriend or a teacher;
- public figures with a(n) (inter)national reputation, such as a politician, a popstar or a sportsman.

Thus, the a-category can be subdivided into two categories: 'acquaintances' and 'public figures'. Table 8 presents the number of texts within this subcategory in relation to age, type of school and sex in proportions together with the total number of texts.

Table 8. Distribution of texts about acquaintances (acq) and public figures (pub) in relation to age, type of school and sex

		girls			boys		
age	type of school	acq	pub	total	acq	pub	total
10	primary	.69	.31	39	.77	.23	30
12	primary	.56	.44	61	.37	.63	51
15	secondary (LBO)	.74	.26	39	.47	.53	36
15	secondary (ULO)	.71	.29	28	.38	.62	34
15/16	secondary (Grammar school)	.86	.14	14	.71	.29	17
Total		.67	.33	181	.50	.50	168

The total proportion of texts per sex shows that girls more often chose to describe an acquaintance (.67 vs .33), whereas the boys wrote about an acquaintance just as often as about a public figure (.50 vs .50). This preference of the girls for writing about an acquaintance can be found in each group ($\chi^2(4)=6.79$, $p=.15$). Boys at the ages of 12 and 15, however, showed an increasing preference for describing public figures when they got older, except for the 15-year-old Grammar school pupils, who more often chose to describe an acquaintance ($\chi^2(4)=17.28$, $p<.01$).

The difference between the sexes is due to the fact that a considerable number of girls - irrespective of age and type of school approximately 25% - did not describe a person they would like to resemble, but one they were or seemed to be in love with.

Consequences for the sample

This preliminary exploration taught us that the original corpus consisted of texts derived from different interpretations of the writers:

a. most of the subjects had interpreted the assignment as 'describe an existing person'. An exception being the grammar school pupils, who more often constructed a person.
b. within the interpretation 'describe an existing person' two topics can be distinguished: an acquaintance from the subject's private neighbourhood and a public figure with a(n) (inter)national reputation. Within these descriptions of acquaintances, a fair number of female subjects chose someone with whom they were in love.

The findings of this exploration had consequences for the selection of the texts for the sample to be used in this study.

The preliminary exploration showed that within the intended interpretation (describing a person one would like to resemble), a distinction had to be made between descriptions of an existing person on the one hand, and of a constructed person on the other. All groups, except for the Grammar school pupils, showed a preference for describing an existing person.

In order to have that preference reflected in the sample, we decided to exclude the texts about a constructed person written by the 10-, 12- and 15-year-old (ULO) subjects. To guarantee, however, a representative sample of the corpus in which also the influence of the topic choice on writing processes could be studied, the texts about a constructed person by the Grammar school pupils were included in this selection.

To be able to make unambiguous comparisons, all sample texts should satisfy the criteria of the assignment. As was stated above in the discussion of table 8, a considerable number of girls had interpreted the assignment quite differently ('describe your adorable boy friend'). In addition, there were texts in which it was impossible to decide whether it was a description about an intended boyfriend or a real ideal. In order to control for this difference in interpretation, we decided to exclude all girls from the selection of the final sample. As the Grammar school pupils describing a constructed ideal person were to be included, and as this group consisted of boys and girls, this group could be used to check for gender differences.[22]

As there was no reason to expect that the group of secondary LBO pupils would yield extra insights into the developmental trends, we decided to take only the ULO-pupils as representatives of the 15 year old ones.

As was stated in section 3.4, the set of texts described above was extended with a sample of 12 texts written by adults.[23] These texts, however, had to be comparable to the others with respect to the topic. This meant that the texts had to be about an existing person. As there was a slight preference for 15-year-old (male) writers to describe a public figure, the adults were asked to describe a public figure they would like to resemble.

The texts that were available for taking the sample are presented in table 9. The sample selected from these texts is described in the next section.

Table 9. Texts qualified for taking a sample in relation to age, type of school and choice of topic

age	type of school	choice of topic			
		existing		constructed	
		acq	pub		N
10	primary	23	7	--	30
12	primary	19	32	--	51
15	secondary (ULO)	13	21	--	34
15/16	secondary (Grammar school)	--	--	34	34
Total		55	60	34	149

3.5.2 Description of the sample

The sample central in this study will be described with respect to subject features and syntactic features in order to check its representativity.

Subject features

From these 149 texts a sample of 86 texts was selected following a procedure to be described below.[24] Since the number of texts about a public figure written by 10-year-old subjects was too small to take a sample from, these texts were not included in the sample.

In the groups of Grammar school pupils and adults both sexes are represented. The group of Grammar school pupils consisted of 12 male and 12 female writers. The adult group added to the sample consisted of 10 female and 2 male writers.

Table 10 shows the sample used in this study classified for age, type of school, and choice of topic. For each group its representativity is given with respect to the 451 texts from which the sample originates (see table 7); the proportion is presented between brackets.

Table 10. The sample of this study in relation to age, type of school, choice of topic and representativity

age	type of school	choice of topic				
		acq	pub	constr	total	representativity
10	primary school	10	--	--	10	(.70)
12	primary school	10	10	--	20	(.98)
15	secondary school (ULO)	10	10	--	20	(.89)
15/16	secondary school (Grammar school)	--	--	24	24	(.44)
21-40	university	--	12	--	12	--
Total		30	32	24	86	--

Syntactic features

The texts in the sample not only differ with respect to characteristics of the writers but also with respect to the text feature 'topic' (an existing or a constructed person). A question is whether the corpus is representative in the sense that it exhibits the developmental differences with respect to syntax that are commonly observed in the literature. The corpus was analyzed for some of the classic criteria for syntax structure.

The indices which will be computed for the corpus are (a) T-unit length, and (b) subordination-index. It will be checked whether syntactic developmental patterns of the selected sample correspond to those found in other studies.

a) The T-unit length

Sentence length is often used to measure language development. Hunt (1965) shows, however, that the length of a sentence is a highly unreliable and inadequate unit for analysis. The most important reason is that 'the sentence' is hard to define, especially in texts written by unexperienced writers. A sentence might be taken as what is written between a capital and a period or other final punctuation mark.

But this causes problems in texts written by novices. Children simply do not always use periods (Hunt 1965: 7). Another reason is that 'the sentence' as a unit is not valid. Sentences can, in principle, be extended by embedding or coordinating clauses. Especially the coordination of clauses by 'and' can lead to an extreme sentence length, which could wrongly be interpreted as a measure of well-developed formulation skill.

Thus, using sentence length as an indicator for development presents practical as well as theoretical problems. In order to overcome these problems Hunt proposed another linguistic unit as a basis for analysis, namely the *minimal terminable unit*, abbreviated as T-unit (Hunt 1965: 21). A *T-unit* is "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it" (Hunt 1970: 4).

Hunt (1965, 1970, 1983) demonstrated that the mean T-unit length is a reliable and valid way to display regular developmental patterns with age.

A lot of researchers have used the T-unit length as a dependent variable in their study of the development of syntactic fluency (e.g. Loban 1976).[25] The T-unit length was also validated by Reesink et al. (1971), who replicated Hunt's study (1970). They conclude that also for the Dutch language T-unit length is a useful instrument to characterize developmental changes.

The results of Hunt's study showed that T-unit length increases with age (1965: 22). As Hunt's results were obtained on the basis of a different task than the one used in this study, a comparison with a study by Van Wijk (1987: 22) might be more informative.[26] In his study subjects performed the same task as in this study. Table 11 presents the T-unit length observed in the studies by Hunt (1965), Loban (1976), and Van Wijk (1987), and those observed in this study.

Table 11. Comparison between studies for observed T-unit length in relation to age

	10	11	12	15	15/16	students
Hunt (1965)	8.60	--	--	11.50	--	14.40
Loban (1976)	8.02	8.76	9.04	10.37	--	13.27
Van Wijk (1987)	--	7.51	8.11	10.12	12.88	--
This study	6.85	--	9.04	9.50	12.78	12.90

The studies by Hunt, Loban and Van Wijk show an increase in T-unit length when writers get older. It should be remarked, however, that the differences in T-unit length over the studies but within an age group may be due to the fact that the T-units are not identically defined in the different studies.

Hunt and Loban did not control for topic and genre. Hunt just took segments up to a maximum of 1000 words per subject. One of the criticisms of his study is that this yields invalid results (Watson 1983: 130). In the study by Van Wijk (1987) topic and genre of the texts were controlled. Besides he used a slightly different definition of the T-unit. The main reason for revising the definition was that the subjects tend to contract sentence initial pronominal subjects. Consequently, (2a) and (2b) would be one T-unit according to Hunt's definition.

To avoid such inflation of the T-unit length, the T-unit was defined as: all constituents organized around the finite verb of the main clause (Van Wijk 1987: 15).[27] According to this new definition (2a) and (2b) are separate T-units, since both have finite verbs.

(2) a He wears glasses
Hij draagt een bril

b and does not yet suffer from loss of hair
en heeft nog geen last van uitvallend haar

The analysis of the texts in this study is based on the adjusted definition by Van Wijk. These results show, an increase, parallel to the results in the other studies, in T-unit length with age from 6.85 to 12.90 ($\chi^2(4)= 35.23$; $p<.0001$).[28]

b) The Subordination-index

The subordination-index is a frequently used instrument to measure syntactic complexity. It shows the proportion of subordinated clauses per T-unit. The index can be computed for free and bound clauses separately (see also Van Wijk 1987: 18). A clause is *free* if its removal does not make the remaining T-unit syntactically or semantically incomplete. For example, (3b) can be deleted without the unit becoming incorrect. A clause is *bound* if its removal makes the remaining T-unit incomplete. See for example (4a), which cannot be removed without creating an illformed sentence.

(3) a She was independent at an early age
Ze was al vroeg zelfstandig

b because she already had her own band at the age of thirteen
omdat ze op haar dertiende al een eigen bandje had.

(4) a That I would like to resemble her
Dat ik op haar zou willen lijken

b would not be completely right.
zou niet helemaal juist zijn.

Bound clauses mostly consist of complements and restrictive relative clauses. Examples of free clauses are adverbial clauses, appositions, and non-restrictive relative clauses.[29]

Table 12 presents the subordination index for the corpus texts, specified for their function: free or bound.

Table 12. Subordination-index for the corpus in this study in relation to age

	10	12	15	15/16	adult
Subordination-index	.22	.31	.38	.89	.86
Free subordination	.20	.25	.29	.57	.46
Bound subordination	.02	.06	.09	.32	.41

Just like T-unit length, the subordination-index shows an increase with age and education ($\chi^2(4)=36.39$, $p<.0001$). Decomposition of the overall index shows that the majority of subordinations are free. Their number increases significantly with age ($\chi^2(4)=17.82$, $p<.0001$). The proportion of bound subordinations is rather small until the age of 15, but increases with a major jump in the texts of the 15/16 year-olds and adults ($\chi^2(4)=40.12$, $p<.0001$).

Concluding, the above investigation of the sample shows that it is not an a-typical one. The syntactically related developmental patterns - with respect to T-unit length and subordination-index - correspond to the ones found in other studies.

The above exploration of the sample shows signs of systematic differences with respect to *shaping* and *encoding* between writers of various ages. The next chapters will focus on the systematic patterns and differences in the conceptual processes *Reflect* and *Select* which leave their traces on a text level. The instrument which is used to transform these traces into comparable units will be described in the next chapter.

Notes

1. *Extrinsic* and *intrinsic* should be looked upon as relative terms. Writers may, for instance, 'whisper aloud' when writing a text, which is of course intrinsic. Pauses, on the other hand, may also arise from extrinsic factors such as saving text in the electronic format of a word processor.

2. Of course, this brief overview is not complete; there are a lot of other techniques. Other on-line extrinsic data and methods are, for instance, composing aloud (as a variant of thinking aloud), 'live' or video observation, and registering brain activities. A variant of text analysis is, for instance, comparing text versions. See also Faigley et al. (1985: 178 ff.) for a short overview of methods to study the composing process, and Bochardt (1983: 6-7) for an overview of methods in relation to specific research questions.

3. For the sake of completeness it has to be remarked that not all interviewing takes place after writing (post-hoc interview). A subject may also be interviewed during writing (concurrent interview; Faigley et al. 1985). This concurrent interviewing is not yet much applied in research on writing. As the term already indicates, the interview takes place simultaneously with the process itself. Therefore it increases the chance of disturbing the process, which could influence the (natural) way of composing.

4. Recent studies in which pause analysis and text analysis are combined - within the same framework as this study - are Sanders, Jansen, Van der Pool, Schilperoord and Van Wijk (1995), Schilperoord & Van der Pool (1995).

5. For a discussion of advantages and restrictions of each level one is referred to Bereiter and Scardamalia (1983: 3 ff.; 1987: 35 ff.).

6. Nevertheless, text analysis is often used as an instrument to quantify texts.

7. See also Cooper: "Though the procedures begin with labelling and enumeration of text features, they can take us well beyond pointing and counting. They provide a (...) linguistics-based description from which we may more confidently interpret differences we find in discourse types and in writers at various stages of development" (1983: 306).

8. They add that the main job is to search for lawfulness "in the protocols". This restriction to protocols seems to indicate that they confuse the method with the level of inquiry, which obviously cannot be meant, given their own statement (1987: 42).

9. Bereiter and Scardamalia treat these levels of inquiry in a sequential and cumulative way. This might symbolize the ideal way of doing research. However, most projects only capture parts of these levels. Taking this interpretation of doing research, it does not necessarily cover all levels. Within the levels a traditional distinction between two different paradigms is hidden: 'descriptive/heuristic' research versus 'testing/theoretical' research. Also simulation may be based on information obtained from levels 3 to 5. Taking this distinction into account, we would propose an alternative for organizing the methods ('levels') of inquiry (see figure A).

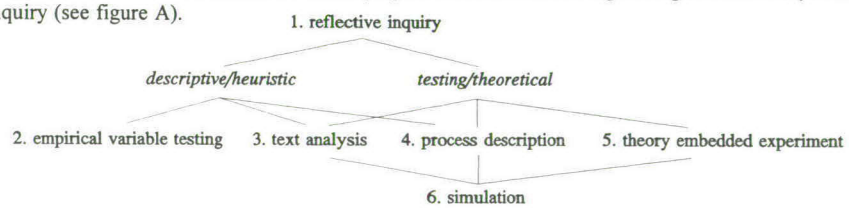


Figure A. Alternative organization of Bereiter & Scardamalia's levels of inquiry

10. We do not claim that text analysis represents or symbolizes human processing; what we aim at is developing a psycholinguistically plausible structure of the text as a result of the analysis.

11. Mann and Thompson specify this as follows: "Plausibility is a threshold concept, based on a degree scale and a conventional way of dividing the scale to provide a binary judgment" (1988: 279).

12. Of course, retrieving and organizing may include hidden operations as well. A text mostly does not display, for instance, what information was supposed to be inappropriate and has therefore been excluded. Nevertheless, the information found in the text leaves no doubt with respect to (final) retrieval and organization.

13. We regard descriptive texts as texts with an additive structure. This addition may be sequential when it describes an event (cf. narratives in Christensen and Virtanen), or it may list-wise when it describes an object (cf. descriptives).

14. This task has been constructed within the framework of a social developmental psychology study (see Lutte, Mönks, Kempen & Sarti, 1969). I am grateful to Prof. Dr. Gerard Kempen for making the corpus available.

15. The original Dutch version was as follows.

    Beschrijf een persoon waarop je zou willen gelijken. Het mag zowel een echt bestaande mens zijn als een ingebeelde mens; of ook nog iemand die in zich de kenmerken verenigt die je in verscheidene personen opgemerkt hebt.
    Vertel iets over die persoon: hoe oud is hij? welk karakter heeft hij? hoe ziet hij er uit? wat is zijn beroep? waar zoekt hij zijn ontspanning en hoe gebruikt hij zijn vrije tijd?
    Zo je het hebt over een echt bestaande persoon, schrijf dan op het einde "bestaande persoon". Je bent niet verplicht zijn naam te vermelden, indien je dit onaangenaam is.

16. With respect to the age of the corpus, we would like to remark that although the texts were not written in the nineties - given the choices of the topics - we assume that their value for studying conceptual processes has not decreased, for these underlying processes are universal and not bound to a particular time with respect to this task (describing a person one would like to resemble). It does not make much difference whether someone described his father in 1965 or does so in 1995. Only those who would like to be like a public figure would probably no longer describe the leading singer of the Beatles but someone like for instance the leading singer of De Dijk. (The differences related to these choices do not relate to Invent-operations but are related to the type and amount of world knowledge a subject has.)
    With respect to the use of the corpus for the study of developmental aspects, one could remark that the influence of the educational system has changed a lot in the last three decades. Nowadays much more attention is paid to functional writing, with the aim to affect Reflect-operations such as determining intentions, and structuring the text with respect to a reader. Such skills would possibly influence free writing tasks as well. Although this might restrict the generalizibility of the results to the current system, comparing the texts in order to reveal developmental changes is still reliable as all subjects (including the adults) were educated in more or less the same time period.
    In addition, we do not want to collect repeated measurements to evaluate the educational system, we only want to detect aspects of conceptual processes.

17. This extension took place in February 1993.

18. LBO stands for 'Lower vocational training', ULO stands for 'secondary modern education'; the Dutch abbreviations are used to refer to the groups.
    The LBO-group consisted of female pupils of the domestic science school (Huishoudschool) and male pupils of the technical school (LTS). Both groups are represented in the tables and the discussion as

'LBO'. The ULO-group consisted of male pupils of the secondary modern school, and will be referred to as 'ULO'.

19. A logistic regression analysis was used in order to check the influence of the variables 'gender' and 'schooltype' in predicting whether a text was to be classified as drop-out or not. The logistic regression model predicted that 83,5% of the texts met the criteria of not being a drop-out. Adding the variables 'type of school' and 'gender' did not yield a better prediction, in other words the percentage of drop-outs is independent of these variables.

20. This interpretation differs clearly from the 'drop-out' texts. In these texts only one constructed person is described, whereas in the drop-out texts more than one person was described.

21. There were no significant differences in the assignment interpretation between the four groups of 10-, 12-, 15-secondary (LBO), and 15-secondary (ULO), nor between the three groups 12-, 15-secondary (LBO), and 15-secondary (ULO).

22. Other studies showed that no differences in language abilities of males and females are apparent, although there do seem to be differences in the way males and females use language (Price & Graves 1980: 151).

23. This group consisted of 2 male and 10 female writers, differing in age from 21-40 years.

24. The texts about an existing person - except for the ones written by the adults - were stratified in terms of length of text. The length of text was set, and on the basis of the median a division was made between short and long texts. In every group five short and five long texts were taken up in the sample.

25. Schuurs (1990: 125) remarks that the T-unit length is to be used as the only measure for determining syntactic maturity. It has, for instance, to be controlled for the number of grammatical errors (Schuurs also mentions some other disturbing variables).

26. Hunt (1965) used a free writing assignment. The subjects were allowed to write about any topic. The only restriction was that it would deliver 1000 words or more.

27. Another advantage of this modification of the T-unit definition was that it removed undesirable differences between English and Dutch (see Van Wijk 1987: 15-16 for examples.)

28. There were no significant differences within the age groups with respect to the choice of topic (existing - acquaintance or public figure - or constructed person); therefore they were lumped together.

29. The subordination-index can also be computed with respect to the position of the clauses. Clauses can take a non-final position, as in (2a), or a final one as in (3b). The scores for non-final and final subordinations are not presented in the tables. The results corresponded to the other indexes: there was an increase with age.

4 PISA: a method to analyze text structure

4.1 Introduction

Text analysis has its history in a literary context. It is used as an instrument to reveal a (deeper) level of meaning in plays, poems, novels, and similar texts. As it developed, text analysis began being applied in psycholinguistic research as an instrument to gain insight into aspects of production and processing, for instance through the analysis of the distribution of given and new information (Chafe 1987; Prince 1981), the gist of a text (Kintsch & Van Dijk 1978), cohesive elements (Halliday & Hasan 1979; Halliday 1994), the structure of narratives (Labov & Waletsky 1967; Propp 1968; Rumelhart 1975; Thorndyke 1977), the structure of argumentations (Toulmin 1958), and the structure of expository prose (Meyer 1975).

In this study, text analysis, and more specifically the analysis of text structure, is used as an instrument to reveal aspects of Reflect and Select as text structure mirrors these conceptual processes (see chapter 3, section 3.3.2). The claim is that once we are able to detect the structure underlying the text, we have a means to induce the strategies and procedures of writers.[1]

This psycholinguistic study of conceptual processes requires an instrument that describes *text structure*. Psycholinguistic and linguistic research have yielded insight into text structure. In this study, the following linguistic theories are of primary importance: Mann and Thompson's Rhetorical Structure Theory (1987; 1988), the Linguistic Discourse Model developed by Polanyi and Scha (1983, 1984; Polanyi 1988), and Van Dijk's superstructure (1986, see also Van Dijk & Kintsch 1983, chapter 7). These methods, however, are not yet sufficiently explicitly defined in detail. In particular, they do not specify in detail how linguistic features indicate the text structure. They leave much to the intuition of the analyst. This decreases the reliability of these methods (see chapter 3, section 3.3.2; Mann & Thompson 1988). One of the most explicit explanations of linguistic connections within texts is the concept of cohesion developed by Halliday and Hasan (1976). Their approach offers a basis for specifying the relation between linguistic features and text structure.

Sanders and Van Wijk have integrated psycholinguistic and linguistic theories and have developed a method called Procedures for Incremental Structure Analysis (PISA; see Sanders 1992, Sanders & Van Wijk 1995a,b).

The further development of this method has led to a number of extensions and reorganisations. This second version will be described in this chapter.

In this chapter the backgrounds and working of PISA, the method used in this study, will be described in more detail (see the marked box in figure 1).

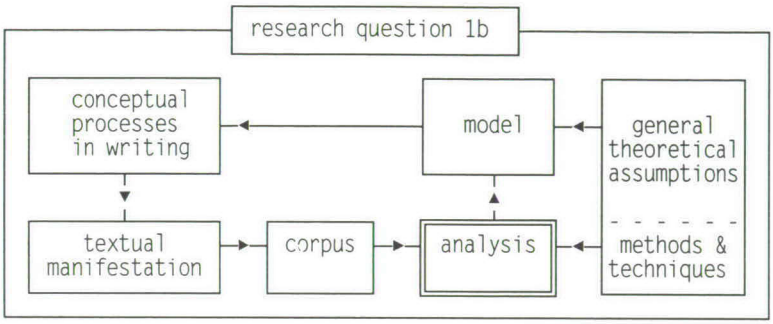


Figure 1. The position of this chapter in the study

It starts with an outline of the theoretical backgrounds of PISA (section 4.2). The way PISA works will be discussed by specifying its output (section 4.3), its input (section 4.4), and its architecture (section 4.5). The chapter closes with an evaluation of the method (section 4.6).

4.2 Theoretical backgrounds

Text structure consists of a relational and a hierarchical aspect (Sanders 1992). The *relational* aspect concerns the meaning of the connections between sentences, for instance reason, exception, or specification. The *hierarchical* aspect concerns the different levels of the connections, and consists of three levels: a micro level, a macro level and a superstructure level (cf. Meyer 1985: 16-20; Van Dijk 1986).[2]

The aim in the present study is to develop a method of analysis that yields a *comprehensive structure*. The output must include the relational meaning between units, as well as the hierarchical position on the three levels. The analysis should be as explicit as possible. This implies that the analyst's intuitions are minimized, that optimal use is made of explicit linguistic elements, and that the use of world knowledge will be made as explicit as possible.

In developing PISA current theories of text structure were used as a foundation to as great

a degree as was possible. Different disciplines have emphasized different aspects of text structure.

Psycholinguistic approaches to text structure tend to focus on one of these aspects. Examples of studies in which the relational meaning of connections is the central issue are Hobbs (1983), and Sanders, Spooren and Noordman (1992). They characterize coherence relations from a cognitive point of view. Examples of studies on hierarchical aspects are Kintsch and Van Dijk (1978), Van Dijk (1986), and Van Dijk and Kintsch (1983). The micro- and macrostructure are, for instance, crucial in the model for comprehending and producing text presented by Kintsch and Van Dijk (1978). The superstructure was studied by among others Van Dijk (1986) and Van Dijk and Kintsch (1983, chap. 7); they specified the superstructure of news texts.

In the *computational framework*, there are two relevant methods which deal with both relational as well as hierarchical aspects. These are the Rhetorical Structure Theory, abbreviated as RST (Mann & Thompson 1988) and the Linguistic Discourse Model, abbreviated as LDM (Polanyi & Scha 1983, 1984; Polanyi 1988).

In the next subsections, different proposals from various research traditions which have contributed to the development of PISA will be discussed. In the order of presentation they are: the cohesion inventory by Halliday and Hasan (section 4.2.1), Polanyi and Scha's LDM, Mann and Thompson's RST (section 4.2.2), and the structure schemes developed by Meyer and by Van Dijk (section 4.2.3). In section 4.2.4, the main characteristics of PISA based on these proposals are summarized.

4.2.1 The cohesion inventory

The most comprehensive account of linguistic features contributing to the connectedness of segments is Halliday and Hasan's concept of cohesion. It provides an inventory of verbal expressions of cohesion. Five types of cohesive ties are distinguished: reference, substitution, ellipsis, conjunction, and lexical cohesion. Halliday and Hasan state that these features indicate a certain semantic relation between sentences that are related to each other.

The main criticism of Halliday and Hasan's cohesion approach is that it is a restricted interpretation of connectedness in texts. Halliday and Hasan only consider explicitly expressed realizations of connectedness (Enkvist 1978: 110; Brown & Yule 1983: 195). The reason why the cohesion approach does not account for the coherence of a text is that implicit relations are ignored. Language users do not completely depend on formal markers of cohesion to identify coherence (Brown & Yule 1983: 198). Sequences of segments without cohesive ties may be quite coherent as is shown in (1) (see also Van Wijk & Sanders 1987; Sanders 1992).

(1) Joost is a good cook. I put on a kilo every week.

This combination of segments is not cohesive, nevertheless it is coherent, as one can easily infer, on the basis of world knowledge, that 'I' have dinner with 'Joost' almost every day, that 'I' appreciate his menus very much and that Joost excels at cooking high-calorie dishes.

On the other hand, it is misleading to assume that where there is cohesion, there is also coherence (Bereiter & Scardamalia 1987: 121). Sequences with cohesive markers may still be incoherent due to a lack of global structure or a shared knowledge domain, as is the case in (2).

(2) Joost likes *to eat* sprouts. *To eat* is *a verb. A verb* is a linguistic expression that indicates an *action* or state. Pulp Fiction is an *action* film worth seeing.

This associative list of segments is cohesive on the basis of lexical reiteration; every segment - except for the first, of course- contains a word used in the preceding segment. The list in (2) cannot, however, be conceived as coherent. A simple overlap of words does not activate global semantic relations and makes little sense.

Clearly, Halliday and Hasan's current list is inadequate and needs extension in three respects:[3]

a. by adding a notion of global structure
b. by enlarging the inventory of cohesive elements
c. by taking into account more divergent aspects within the domain of cohesive features.

a) *Adding a global structure*

Segments are not only connected on a micro level, they may cohere on a macro and top level as well. This can be illustrated by the fragment presented in (3a).

(3a)
1 Joost weekly goes to the market.
2 It is held every Saturday.
3 You can buy almost everything there.
4a In order to please his girlfriend,
4 he usually buys flowers,
5 or surprises her with sweets.
6 Last Saturday he bought ten fresh treacle waffles.
7 He ate a warm one at once.
8a When he came home,
8 he discovered an empty sack in his shopping bag.

In terms of cohesiveness, the ties that constitute the cohesion in this fragment are of different types: reference (*Joost*, *he*), lexical cohesion (*buy-buys-bought*, *sweets-treacle waffles*), substitution (*a warm one*), ellipsis (the subject *he* in segment 5), and conjunctions (*in order to*, *or*, *when*). In (3b), these ties are marked; the references are italicized, the other cohesive ties are underlined.

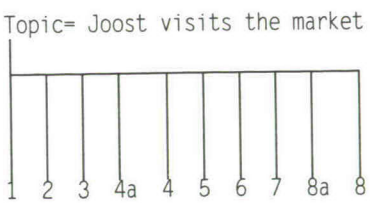
(3b)
1 Joost weekly goes to the market.
2 *It* is held every Saturday.
3 You can buy almost everything *there*.
4a *In order to* please *his girlfriend*
4 *he* usually buys flowers,
5 or surprises *her* with sweets.
6 Last Saturday *he* bought ten fresh treacle waffles.
7 *He* ate a warm one at once.
8a When *he* came home,
8 *he* discovered an empty sack in *his* shopping bag.

Each segment contains at least one expression referring backwards to a previous one. In the first segment two concepts are introduced: *Joost* and *the market*. These two concepts play an anchoring role in constituting connectedness. The second and third segments include a reference to *the market*; from 4 onwards, the segments have a reference to *Joost*. In addition, segment 4a introduces a new concept: *his girlfriend*. She is referred to only once in segment 5 by the pronoun *her*. The sweets that are introduced in 5 are specified in 6 as *treacle waffles*. One of these waffles is further referred to in segment 7 by the expression *a warm one*.

Moreover, the segments are connected by lexical cohesion (that is *buy* in segment 3, *buys* in 4, and *bought* in 6; *sweets* in segment 5, *treacle* waffles in 6), substitution (a warm *one* in segment 7), ellipsis (in segment 5, *he* is contracted), and conjunctions (*in order to* in segment 4a, *or* in 5, *when* in 8a).

On the basis of these ties, it could be argued that fragment (3) consists of a simple string of segments which are hierarchically equal, as is depicted in (4).

(4)



Using only the above mentioned cohesive ties, all of the segments would be linked to the first one, which states the topic 'Joost visits the market', on the basis of argument overlap. Every segment contains a tie referring to either *Joost* or *the market* introduced in segment 1. Segment 2 is linked to the first segment on the basis of the reference to *the market* mentioned in 1 (*it*). Segment 3 has a reference to the market in the form of *there* which makes a connection to the first segment. The segments 4 to 8 all comprise a reference to *Joost*, by which they are connected to segment 1.

This description, however, is not a good account of the structure. It suggests that the fragment is a simple enumeration of segments, without any further coherence. This is obviously not the case. There is more coherence between the segments than just the individual relation of each segment to the first one. The presentation in (4) ignores the relations on the macro- and superstructure level.

With respect to the macro level relations, fragment (3) can be differentiated into three groups of segments: a *descriptive* pattern, a *response* pattern and a *narrative* pattern.

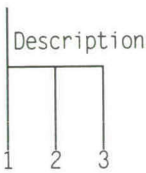
After the introduction of *Joost* and *the market* in the first segment, the fragment continues with two segments which *describe* characteristics of the market, it is held weekly (S2), and there is a wide range of products (S3) (cf. Meyer 1985). Segment 2 starts with a reference to it; segment 3 contains a reference to *the market* which is not the subject; it has the postverb position. Although this anaphor does not take the pre-verb position, the segment elaborates on *the market* as the subject (*you*) is impersonal.

Segment 4a states a *goal* that Joost wants to reach (*in order to please* his girlfriend (S4a)) and the *instruments* he is using (*buy flowers* (S4), or *sweets* (S5)). The goal is realized in segments 4 and 5 by specifying the instruments (cf. Grimes 1975; Meyer 1985; Sanders 1992; Sanders & Van Wijk 1995a).

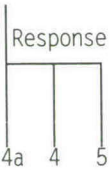
The three final segments, the anecdote in S6 to S8, *narrate* an event (cf. Meyer 1985 [4]; Van Dijk 1986). In these segments Joost's activities are described, triggered by the type of verbs (action verbs instead of state verbs) that are used, their temporal form (past tense), and the presence of sequence markers (*last Saturday*, *when*). The action verbs in the past tense indicate a sequence of activities: he *bought*, *ate* a waffle, *came* home, and *discovered* an empty sack. This sequence is supported by the adverbial expression *last Saturday*, which marks a specific point and opens a temporal narrative line, and by the adverb *when* which indicates another temporal reference point.[5] It marks the next phase in the sequence of events without making explicit all of the previous ones, that is, the crux of the anecdote that Joost imperceptibly also ate the other nine waffles.

The hierarchical structures of these relations on a macro level are represented in (5a), (5b), and (5c), respectively; segments 1 to 3 form a Description pattern, 4a to 5 a Response pattern, and 6 to 8 a Narration pattern.

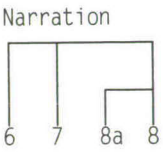
(5a)



(5b)



(5c)



These patterns are related to each other on a superstructure level. In the descriptive pattern, depicted in (5a), the topic, *Joost visits the market*, is introduced with a small elaboration of two main characteristics of the market. The latter of these two, segment 3, is specified by the response pattern, depicted in (5b). Thus, the response pattern is hierarchically subordinated to segment 3. Furthermore the narrative pattern elaborates Joost's usual behaviour mentioned in segment 4, which is part of the response pattern. This elaboration makes the narration pattern subordinate to segment 5. Thus, the text is structured along the lines of a general statement (description pattern), followed by a specification of one aspect of that statement (response pattern), and finished with an illustration of an element mentioned in the specification. The entire hierarchical structure, which takes into account the coherence on the macro and superstructure levels, is depicted in (6).

(6)

b) *Extending the domain of cohesion by other features*

Halliday and Hasan identify five categories of lexical forms and phrases. It remains unclear why they only use these specific categories and why they leave aside other cohesive markers, such as *type of verb* (action verbs versus state verbs), their *temporal state* (present tense versus

past tense), and the specific functions of *subordinated clauses*.

The type of verb and its temporal state are informative about the text structure. Action verbs in the past tense, for example, such as *bought*, *ate* and *discovered* in (3), determine the global relations between the segments for instance. They signal that these segments represent a narrative.

Subordinate clauses have different effects on the text structure, dependent on their position and function. If they follow the main clause, they generally consist of background information, which makes them subordinate to the main clause. However, if they precede the main clause, as is the case in segments 4a and 8a in fragment (3); they may consist of foreground information indicating a goal or a problem, as is the case in segment 4a. These initial subordinate clauses stating a goal or problem dominate the main clause (see also Sanders & Van Wijk 1995a and De Bruijn 1995, who discuss this phenomenon of response patterns in detail[6]). Initial subordinate clauses which do not state a response provide fronted background information, and become subordinate to their main clause, as is the case in segment 8a in fragment 3.[7]

c) *Extension of the inspected features within the domain of cohesion*

The cohesive features of Halliday and Hasan's inventory are not examined to the fullest extent. Halliday and Hasan describe the cohesion of texts in terms of the *type* of connective tie and the *distance* 'separating the presupposing from the presupposed' (see 1976: 329 et seq.). These two criteria are not sufficient to distinguish connective ties which are equal in type and distance, see, for example (7).

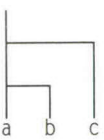
(7) a. *Joost* was wearing *a new shirt* at the party.
b. *It* fitted *him* perfectly.
c. *He* looked very smart in *it*.

The cohesive ties in (7b) and (7c) are of the same type, both referential, and have the same distance to their antecedents, which is the segment directly preceding them. Thus, the type of reference and the distance to an antecedent are insufficient for determining the hierarchical positions of these segments. The *position* of the anaphors with respect to the verb, and their *status* with respect to the topic of the text also have to be taken into account. If a segment contains two referential expressions, anaphors in the *preverb* position functioning as subject determine the relation and overrule anaphors in the postverb position. An illustration of this is 7b which elaborates on *new shirt* and not on *Joost* as *it* takes the preverb position and overrules the cohesive strength of *him*. The same holds for segment 7c; the anaphor in the

preverb position, *he*, determines that the segment elaborates on Joost and not on the shirt. Linking 7c, however, is a little more complex as Joost is referred to in 7b, as well as in 7a. In order to decide which of these two sentences 7c should be linked, one has also to take the status of the anaphor into account. The two anaphors in preverb position, *he* and *it*, differ with respect to the status of the concepts they refer to. The difference between these two is that *he* in 7c refers to the protagonist of the fragment, and *it* in 7b elaborates on an attribute of this protagonist (the shirt). As 7c elaborates on the protagonist, it should be linked to the segment in which that protagonist is, likewise, in the preverb position, thus, 7a.

These aspects of the anaphors form the condition for determining the hierarchical position of segments; they are depicted in (8).

(8)



In conclusion, it is insufficient to take into account only the type of connective tie - for example an anaphoric expression - and its distance to the antecedent; it is also necessary to know its position within the sentence and its status in order to link the segments in the correct fashion. Thus, within the domain of anaphoric ties, the aspects which acts as triggers should be extended to include the position and the status of the tied elements.

Despite these critical remarks, Halliday and Hasan's inventory of linguistic features still offers a very useful starting point for describing text structure. In PISA, this inventory is extended in the three directions mentioned above, by using research which has been done since 1976 (examples of such studies with respect to anaphors and references are Bosch 1983, Fox 1987, Clifton & Ferreira 1987; with respect to connectors Martin 1992; and with respect to sentence form Givón 1979, Ramsay 1987, Verhagen 1992).

4.2.2 RST and LDM

Two relevant methods for describing text structures, the Rhetorical Structure Theory (Mann & Thompson, 1987, 1988), and the Linguistic Discourse Model (Polanyi 1988; Polanyi & Scha 1983; 1984) originated in the field of computational linguistics. Both approaches yield the structure of a complete text, though in quite different ways.

Rhetorical Structure Theory

The RST yields a hierarchical structure with a specification of the relational meaning of each connection. It decomposes a text into either terminal segments or sets of segments which require further analysis.[8] The rhetorical meanings of the relations between the segments are assigned on the basis of an open set of relations which might be reasonable constructs in a theory of text structure (Mann & Thompson 1988: 249). The relations are defined in a more or less exact manner. For each relation, the constraints on the segments and the effect of the relation are specified (Mann & Thompson 1988: see appendix). Assigning these rhetorical relations, however, takes place by inspecting the relation between clauses (bottom-up), as well as the complete text (top-down). As there are no exact rules for switching between bottom-up and top-down analysis, assigning these rhetorical meanings still leaves much to the analyst's intuitions (see Sanders & Van Wijk 1995a). As in RST, PISA assigns rhetorical meanings to the relations.[9]

Linguistic Discourse Model

The LDM provides an explicit formal treatment of discourse phenomena and is formalized as a discourse parser which builds up the structural description of a discourse on a left-to- right, clause by clause basis (Polanyi 1988: 602-603). Discourse is processed *incrementally*. Thus, LDM does not proceed top-down, but bottom-up. The meanings of the interclausal relations are not precisely specified.

In building a discourse parse tree, the segments, which are called discourse constituent units (dcu's)[10]- are either coordinated or subordinated. Sequential dcu's are considered *coordinate* to each other, whereas dcu's expanding a proposition in an immediately preceding clause are treated as *subordinated* to the clause they expand. "Ultimately, the decision of whether to subordinate or coordinate a given unit must be made using real world knowledge and inferential procedures." (1988: 611)[11].

As is also true from LDM, PISA operates incrementally. It does not inspect the entire text first but builds a structure from left-to-right, clause by clause. The manner on which a segment is linked does not depend on what follows. In addition, the hierarchical positions PISA assigns in the microstructure, the subordinative and coordinative, correspond to those in the LDM.

In contrast to the LDM, PISA aims at circumventing the unrestricted use of world knowledge and inferential procedures by operating *procedurally*. In order to minimize the use of the analyst's intuition, the rules for text analysis are made as explicit as possible. This is realized by using production rules (in the form of condition-action pairs, if-then statements; see the appendix of this chapter). This procedural character ensures the intersubjectivity of the analysis. In addition, the use of knowledge is confined. This is operationalized by making

knowledge used explicit in a knowledge base. This base presents the information which is relevant to the topic of the text and the information which is specifically needed to assign the macrostructure and superstructure (see section 4.4).

4.2.3 Structure schemes

Psycholinguistic ideas about conceptual patterns guiding text production (Bereiter & Scardamalia 1987), and about overall text structure (Van Dijk 1986; Van Dijk & Kintsch 1983; Meyer 1975, 1985) are important in analyzing text structure. In text production, a writer makes use of 'Discourse Schemes', schemes which specify the kinds of elements to be included in the discourse and to a certain degree describe their arrangement (Bereiter & Scardamalia 1987: 7). They are used as guiding principles in writing.[12] These discourse schemes do not cover the entire text, but only groups of segments. They should not be confused with superstructure, which orders textual sequences of sentences and assigns specific functions to such sequences (Van Dijk 1986: 158).[13]

Superstructures are conventional forms that characterize a specific discourse genre. They order textual sequences of sentences and assign specific functions to such sequences. They organize higher level units (Van Dijk 1986: 158). Examples of research with respect to this level are the story grammar (Rumelhart 1975, Wilensky 1983), and Van Dijk's superstructure for news schemata (Van Dijk 1986; Van Dijk & Kintsch 1983).

In accordance with these theories, PISA assigns a structure to the macro level and to superstructure level.

4.2.4 A summary of the characteristics of PISA

In this section, the characteristics of PISA, which is primarily based on the systems discussed above, are summarized.

PI*S*A is a an instrument for *S*tructure *A*nalysis designed to yield a labelled hierarchy of the text structure on the basis of explicit linguistic features. Its output is a comprehensive text structure which consists of both relational aspects and hierarchical aspects on a micro-, macro- and superstructural level.[14] In assigning the structure, PISA inspects text segments for the presence of explicit indicators along the lines of an extended inventory of linguistic features contributing to text structure.

*PI*SA is *P*rocedural and operates *I*ncrementally. Text segments are inspected for the presence of coherence indicators through the application of explicitly formulated rules. PISA does not first inspect the entire text, but instead it builds a structure incrementally. Although no claims are made on the psycholinguistic validity of PISA's on-line application, its (off-line) result should be acceptable in the sense that it resembles the mental representation of a writer.

In order to explain how PISA operates, its output (section 4.3), its input (section 4.4), and its architecture (section 4.5) are described. This is illustrated by the example texts presented in tables 1 and 2. The first text was written by an adult (see table 1) and the second by a 10-year-old boy (see table 2).

The texts are split into segments. These segments correspond to clauses with the exception of restrictive relative clauses, clausal objects and clausal complements. The latter are considered to be part of their host clauses rather than separate clauses (cf. Mann & Thompson 1988: 248). In addition, three subclausal structures are considered to be segments (Sanders 1992: 115; Sanders & Van Wijk 1995a): (1) a second conjunct of a coordinated clause in which only one major constituent is contracted, (2) a non-restrictive apposition and an infinitive clause, and (3) major constituents which rephrase the directly preceding segment explicitly indicated by a marker ('such as', 'for instance').

Table 1. Text 1 (written by a young adult, university level)

1	I would,
1a	if forced to make a choice
1	like to be like Ed Nijpels.
2	He looks quite nice,
3	is well-dressed,
4	and expresses himself very well.
5	Besides, he has undertaken all kinds of activities alongside his rather varied career.
6	It is this variation that appeals to me very much.
7	His "main job" at this moment is being mayor of Breda,
8	but as a compere on TV he holds his own as well!
9a	When he relaxes after his busy work,
9	he can be found on the tennis court
10	and the evenings he likes to be with friends or just on his own reading a nice book!
11	Because of his open character he makes friends easily
12	and makes contacts easily.
13	As a mayor he tries to stand close to the people.
14	The age of Ed Nijpels
14a	(round about 40)
14	is an age which has many advantages:
15	one has had the chance (and the time!) to make something of one's life.
16	one does not need to be a 'he-man' any more.
17	Summarizing, I must say that in any case it looks as if Ed Nijpels feels comfortable
18	or
18a	to put it informally
18	'he likes who he is'
19	and thàt is something worth striving for!

Table 2. Text 2 (written by a 10-year old subject, primary school level, includes spelling errors)

1	He is a teacher in Heervarebeek.
2	He teaches children to learn
2a	like arithmetic Language History Geogaphy.
3	He is round about 40 years old.
4	He wears a suit and a white shirt, and black shoes and looks healthy.
5	Children teaching.
6	On Sundays, Wednesday Afternoons, and Saturday Afternoons cycling
6a	Correcting Notebooks
6b	Walking in the woods.
7	His naame is Uncle Harie.
8	Zometimes he goes to tilburg To the families
8a	and window-shopping on Sundays as well.

4.3 The output of PISA

A PISA-analysis yields as output a labelled hierarchy, which consists of four subparts: a microstructure, a macrostructure, a superstructure, and the relational meanings of the linkings. This output is depicted in figure 2.

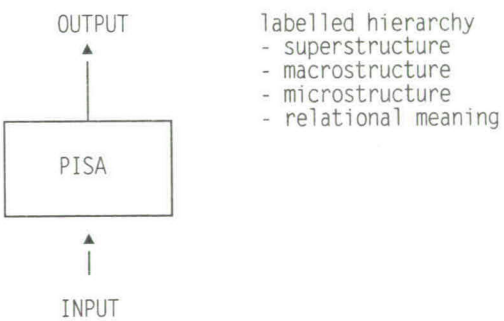


Figure 2. The output of PISA

The output of PISA is a structure graph built from the segments and corresponding to their order in the text.[15] If PISA is applied to the two example texts, it yields the output presented in figures 3 and 4.

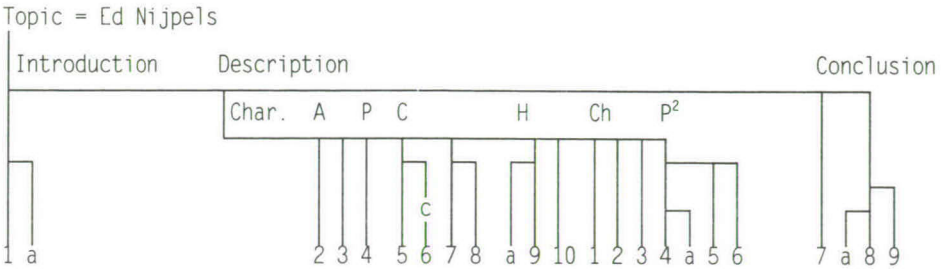


Figure 3. Structure graph of text 1 presented in table 1

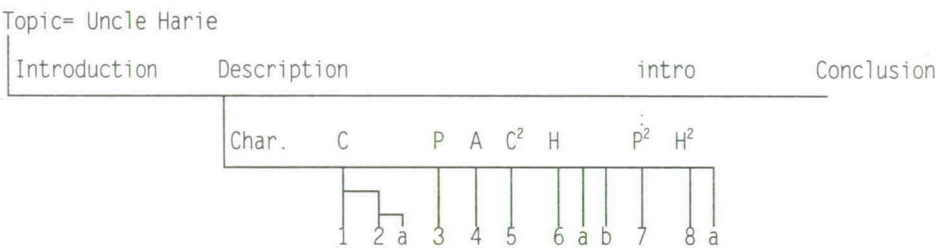


Figure 4. Structure graph of text 2 presented in table 2

Superstructure

The top layer of these structures depicts the global organization of the text. This *superstructure* represents the functions of the segments on an overall level. They correspond to the main parts of the text, each having a specific function. The superstructure of descriptive texts in which a person is characterized consists of a tripartite organization: an introduction, a description and a conclusion (see figure 5).[16]

Figure 5. Superstructure of a descriptive text

The global organization of text 1, depicted in figure 3, consists of an *introduction* (*I would like to resemble Ed Nijpels*), followed by the actual *description* (a list of characteristics), and rounded off with a *conclusion* (a summary). The superstructure of text 2 about Uncle Harie, depicted in figure 4, only comprises a descriptive part; the conclusion is missing, the introduction enters the description.

As the superstructure represents the *function* of (groups of) segments, one would expect that the central part of a text would consist of a factual enumeration of characteristics. However, this is not always the case. Characterizing a person one would like to resemble may lead to evaluative remarks in the form of comment. Two types of comments can be distinguished: explicitly marked statements and implicit ones. An example of an explicitly stated comment is segment 6 in the first text (*it is this variation which very much attracts me*). This segment indicates an evaluation of the career described in the previous segment. It includes an explicit qualification of Ed Nijpels's career. The writer gives his opinion about one specific characteristic of the topic. As segments including these comment statements are not descriptive in the strict sense of the word, they need to be distinguished from the purely descriptive ones; they are all marked with a 'c' in the structure graph.

The texts also contain implicit comments, such as segment 2 in the second text (*he looks quite nice*). This segment is a qualification of the protagonist's appearance. As PISA, however, is dependent on explicit indicators, such as deictic elements referring to the writer, these implicit qualifications are not considered evaluative; they remain unmarked in the structure graph. Only those qualifications which include an explicitly marked opinion are marked in the structure graph.

Macrostructure

The descriptive section ('Description' in figures 3 and 4) within this superstructure consists of a list of characteristics, represented by the 'Characteristic-line' (referred to as 'Char.' in the structure graphs).[17] This is not an undifferentiated list of features attributed to the protagonist. The description of a person consists of several themes (Van Wijk 1992). The themes in the 'Description'-section correspond to those mentioned in the assignment for writing the corpus text: personal particulars, character, appearance, career, and hobbies (see also chapter 3, section 3.4). The themes are abbreviated on the characteristic lines (Char.): Personal particulars (abbreviated to P), Character (Ch), Appearance (A), Career (C) and Hobby (H).

This thematic specification of the descriptive segment is called the *macrostructure*. It defines the *themes* within the descriptive part of a text. An example of a macrostructure in the description of persons is depicted in figure 6. The actual order of the themes may vary per text, and some themes may be discussed more than once in the same text.

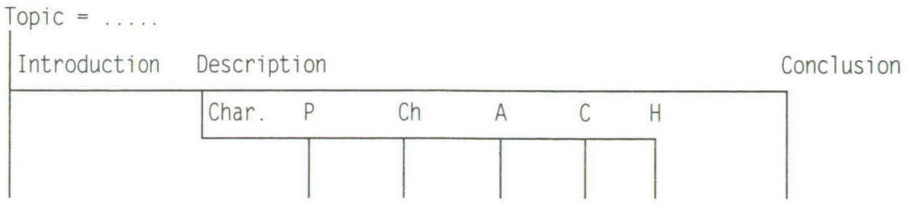


Figure 6. A macrostructure of an arbitrary description of a person

The 'Description'-section of the first text shows a list of Ed Nijpels's characteristics, represented on the 'Characteristics line' (see figure 3). The macrostructure shows that the writer discusses, in that order, Nijpels's appearance (indicated by 'A'), his personal particulars ('P'), his career ('C'), his hobbies ('H'), his character ('Ch'), and again his personal particulars ('P').[18] In this list, the theme 'personal particulars' is discussed twice. The second discussion is indicated by a superscript '2' added to the abbreviation of the theme.

Microstructure

Within these global parts of the texts, segments are also connected hierarchically on a micro level. There are two ways in which individual segments can be linked:

- a segment can be *coordinated* to a preceding one, such as segments 10 and 11 in figure 3 and segments 3-4-5 in figure 4.
- a segment can be *subordinated* to a preceding one, see figure 3 in which 6 is subordinated to 5, and 15 is subordinated to 14. See also figure 4 in which 2a is subordinated to 2.

Relational meaning

Each connection has a certain meaning, such as evidence, reason, or contrast. These relational meanings are not restricted to one particular level. They play a role on every hierarchical level of the structure. For practical reasons, the relations are written down in a separate table.[19] The relational meanings of the connections in text 1 are presented in table 3, those of text 2 in table 4.

Table 3. Relational meaning of the connections in text 1

Segment	Relation
1	SPECIFICATION (Topic)
1a	CONDITION (1)
2	SPECIFICATION (1)
3	LIST Specification (1)
4	LIST Specification (1)
5	LIST Specification (1)
6	EVALUATION (5)
7	LIST Specification (1)
8	CONTRAST (7)
9a	BACKGROUND (9)
9	LIST Specification (1)
10	LIST Specification (9)
11	LIST Specification (9)
12	LIST Specification (1)
13	LIST Specification (1)
14	LIST Specification (1)
14a	LIST Specification (14)
15	LIST Specification (14)
16	LIST Specification (14)
17	CONCLUSION (1)
18a	ALTERNATIVE (17)
18	CONCLUSION (1)
19	EVALUATION (18)

Most segments specify characteristics of the topic, for example segments 2, 3 and 4. But there are also different meanings of the relations between the segments. Segment 1 (*I would like to resemble Ed Nijpels*), for example, is only true, given the condition mentioned in 1a (*if there is no way out*). Another example is the relation between 9a and 9; segment 9a (*When he relaxes after his busy work*) forms the background for segment 9 (*he can be found on the tennis court*). Likewise, the segments 5 and 6 are illustrative in this regard; segment 5 (*...next to his rather varied career*) is evaluated in 6 (*It is this variation which very much attracts me*).

Table 4. Relational meaning of the connections in text 2

Segment	Relation
1	SPECIFICATION (unknown Topic)
2	CONSEQUENCE (1/teaches)
2a	LIST Specification (2)
3	LIST Specification (unknown Topic)
4	LIST Specification (unknown Topic)
5	LIST Specification (unknown Topic)
6	LIST Specification (unknown Topic)
6a	LIST Specification (unknown Topic)
6b	LIST Specification (unknown Topic)
7	LIST Specification (Topic)
8	LIST Specification (Topic)
8a	LIST Specification (Topic)

The second text mainly consists of specifications of characteristics of Uncle Harie, with the exception at segment 2, which is a consequence of 1, and segment 2a which specifies 2. Note that the identification of the protagonist is not presented until segment 7.

The structures presented above form the output of PISA. The question is, of course, how this labelled hierarchy is generated. In order to answer that question a closer look must be taken at the *input* of PISA and its *architecture*.

4.4 The Input of PISA

The input of PISA consists of parsed strings and a knowledge base which includes both discourse knowledge and world knowledge. This input is represented in figure 7.

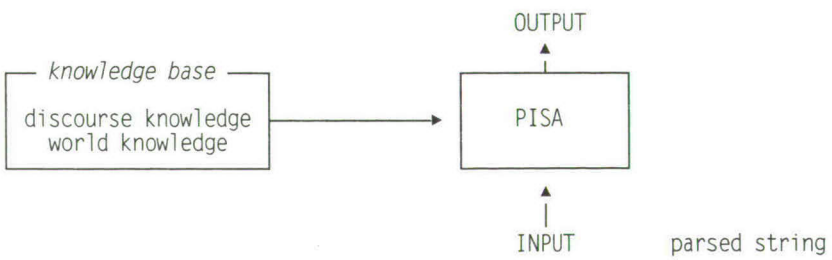


Figure 7. Specification of PISA's input

The procedure accepts as input a *parsed string*, which consists of partitioned text segments which are lexically and syntactically parsed (see Sanders & Van Wijk 1995a). This lexical and syntactical parsing can be achieved using a simple parser working on a left-to-right, word-after-word basis (e.g., Tomita 1986).[20]

The knowledge PISA uses in analyzing the segments consists of *discourse knowledge* and *world knowledge*, both stored in a knowledge base. This knowledge is made explicit to as great a degree as this is possible, as the application of unspecified knowledge to a text may lead to different results between analysts and may decrease the reliability of the analytical instrument.[21]

Discourse knowledge

On the basis of discourse knowledge it can be revealed how texts are structured. The discourse knowledge made available to PISA consists of two structure schemes: one representing the *superstructure*, and one representing the *macrostructure* (see section 4.2).

A superstructure for a descriptive text consists of an introduction, followed by the actual description of the protagonist's characteristics, and a conclusion in the form of a summary statement (cf. Van Dijk 1986; see also section 4.3).[22]

The macro structure depicts the themes discussed in the central part of a descriptive text. The themes of the macro structure, relevant in this study, are based on the suggestions in the assignment (personal particulars, character, appearance, career, and hobbies; see also 4.3 and chapter 3, section 3.4).

World knowledge

The world knowledge PISA uses is operationalized as a database in the form of encyclopedical knowledge as is presented in table 5.[23] This knowledge base is primarily needed for linking the segments to the themes in the macrostructure; its use is restricted as much as possible.

Table 5. Part of the knowledge base on a person's characteristics

Item	Theme in macrostructure
to play tennis	Hobby, unless there is an explicit reference in the segment to another theme (e.g. Career)
to make contacts easily	Character
(to be) well-dressed	Appearance
(to be a) mayor	Career
(to be) 40 years old	Personal particulars

The expression 'to play tennis', for instance, indicates that a segment relates to the theme Hobby, unless it is explicitly mentioned that playing tennis is someone's profession. In that case, the segment would belong to the theme 'Career'. Other examples are 'to make contacts easily' which refers to the theme Character, 'being well-dressed' to Appearance, 'being a mayor' which is a profession that refers to the theme Career, and finally, age, '40 years' old which is related to the theme 'Personal particulars'. Apart from 'real personal particulars' this theme comprises a rather heterogeneous set of features. It is a miscellaneous theme as it includes characteristics that are not intrinsically related to one of the other themes, such as specific talents (playing the piano well), idiosyncrasies (biting nails) or personal belongings (owing a telescope). The lexical form of these items may vary from words to verbal expressions. The information labelled as 'theme' consists of declarative knowledge. This input is transformed by PISA into output consisting of a labelled hierarchy. The crucial question is how exactly does PISA operate? This question is answered in the next section, which deals with the architecture of PISA.

4.5 The architecture of PISA

PISA consists of three components called INSPECT, CONNECT and INTEGRATE. They are schematized in figure 8.[24]

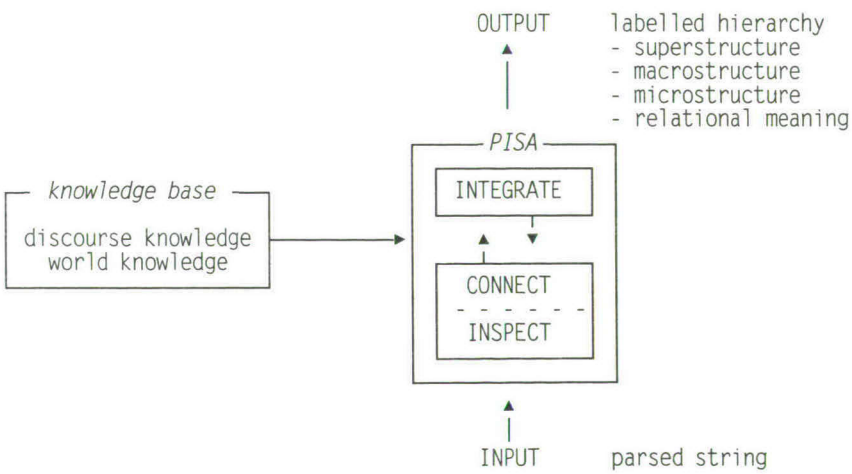


Figure 8. The architecture of PISA

The INSPECT component examines the parsed strings for the presence of linguistic elements which suggest connections to the micro-, macro- and superstructure, and which further suggest relational meanings. It passes on a list of inspected elements and their relevant features (type, token and position in the segment) to CONNECT and INTEGRATE.

CONNECT is a central processor which decides on the microstructure of segments (subordination or coordination), and the meaning of these linkages (such as cause or specification). It operates by means of condition-action pairs. The INSPECT results are matched with the conditions of the production rules. The first condition that is true results in a bipartite action: the assignment of a relational meaning and a hierarchical position in the microstructure. The other conditions which are true result in suggestions for alternative actions.

INTEGRATE enriches the CONNECT decisions with world and discourse knowledge and decides on the position in the macro- and superstructure and the associated relational meanings.[25] For each segment, it is decided whether a new theme is being started in the macrostructure. In addition, INTEGRATE checks the linkings in the structures this far constructed for typical or isolated connections. On the basis of the suggestions for alternative linkings it may decide to replace the linking, with a more adequate one.

INSPECT

INSPECT examines segments for linguistic elements indicating connections to the micro-, macro- and superstructure, and the relational meanings of the connections. The elements inspected belong to three categories: *contraction* (cf. Halliday & Hasan's ellipsis), *connector* (cf. Halliday & Hasan's conjunction), and *referent* (cf. Halliday & Hasan's reference and substitution). INSPECT not only examines segments for the presence of these elements but also for their type, token and position, and, in the case of a referent, INSPECT also searches for an antecedent.[26]

Halliday and Hasan's category denoting lexical cohesion is not included since it is rather unspecified (Stotsky 1983). It is the cohesion that arises from semantic relationships between words and is denoted as reiteration and collocation.[27] However, there is no clear and consistent principle underlying this kind of cohesion. It is primarily based on lexical knowledge related to the topic. With respect to the inspection of linguistic elements indicating a relation to the micro-, and the macro- as well as the superstructure, this category has been filled with markers responsible for these levels within the text structure. These markers indicate an action, a modality, a temporal aspect, a response, a theme, and a relation to the superstructure.[28]

The inventory of linguistic elements inspected by PISA is presented in table 6.

Table 6. Inventory of linguistic elements examined by INSPECT

Coherence level	Manifestation	Linguistic element	Selected references
micro	direct	1. contraction	Halliday & Hasan 1976 (ellipsis)
		2. connector	Halliday & Hasan 1976 (conjunction)
		3. referent	Halliday & Hasan 1976 (reference, substitution)
		4. action marker	Sanders 1992; Sanders & Van Wijk 1995a
		5. modal markers	Sanders 1992; Sanders & Van Wijk 1995a
		6. temporal marker	Sanders 1992; Sanders & Van Wijk 1995a
	derived	7. response marker	Sanders 1992; Sanders & Van Wijk 1995a
macro	direct/derived	8. theme marker	this study
super	direct/derived	9. superstructure marker	this study

This inventory is framed along two dimensions:

a. the level at which the linguistic elements establish coherence: the micro-, the macro-, and the superstructure level.

b. the manner in which coherence is manifested. Elements vary in the manner in which they contribute to coherence. A linguistic feature may indicate coherence *directly*, or in a *derived* form which demands (extra) interpretation on the part of the analyst. This distinction between direct and derived meaning corresponds to two ends of a continuum of interpretations of lexical meanings by means of world knowledge. It can be illustrated using the theme markers in the macrostructure. The phrase '*his hobby* is playing badminton' contains the lexical item 'hobby', which directly indicates a relation to the theme 'hobby'. In a phrase like '*after his work* he plays badminton' the theme in the macrostructure has to be derived on the basis of the present expressions.[29] The expression '*after his work*' indicates a period of time in which the person described has time off. As the activities in an individual's time off are classified as hobbies on the basis of world knowledge, unless indicated otherwise, this segment indicates a relationship to the theme 'hobby' but only in a derived manner. Deriving a meaning takes place on the basis of (explicit) world knowledge and the preceding text.

Following the cohesion inventory, INSPECT examines segments for contracted subject and/or verbs (*contract*), the presence of connectors (*connector*), and nominal or pronominal referents (*referent*). Furthermore, this main function inspects segments for the presence of markers relevant to the microstructure, such as:

- *action markers*, main action verbs as opposed to state verbs (Vendler 1967);
- *modal markers*, modal auxiliary verbs;

- *temporal markers*, lexical items (adverbs or prepositions) or phrases denoting a temporal aspect, or verb-inflexion (temporal connectors are inspected by *connector*);
- *response markers*, lexical items or phrases denoting a problem or goal.

With respect to the macro- and superstructure INSPECT searches for:
- *theme markers*, lexical items or phrases indicating a relation to the theme in the macrostructure (this study for lexical expressions; see Redeker 1991 for push and pop markers);
- *superstructure marker*, lexical items or phrases indicating a relation to the superstructure (Redeker 1991; Van Dijk 1986).

INSPECT examines each segment for the presence of these linguistic elements. If an element is present its *type*, its *token* and its *position* within the segment is inspected.

First, INSPECT examines whether an element is present. If this is the case, the type and the token of the element are inspected. If, for instance, a connector is present, it is noted in which position the connector is (within a main clause, an initial subclause, or a final subclause) and which specific type it is (for example, causal, additive, temporal). With respect to the position of elements, three different positions are distinguished: an element can precede the main verb (preverb position), it can follow it (postverb), and verbs of course take the verb-position. The inspection of the position of elements takes place on the basis of the word order in English (see Sanders 1992).[30]

After examining a segment for the above-mentioned elements and their features, INSPECT passes on a list to CONNECT with the inspection results of elements indicating coherence on the micro level, and it passes on a list to INTEGRATE with the inspection results of elements indicating coherence on the macro- and superstructure level.

The inspection of segments is illustrated in detail by the two text fragments which are presented in (9) and (10); inspection of a entire text is illustrated using text 1 and is presented in table 1. Fragment (9) is part of a description of the tennis player André Agassi, fragment (10) is part of a description of the Dutch football player E.P. 's Graafland.

(9) '*because he belongs to the top players,*
he makes a decent way of living '

(10) '*If the players are quarrelling*
he interrupts
and it stops right away'

Inspection of the clauses in italics yields the results presented in table 7 and table 8. If a feature is present (indicated by 'yes' in the column 'present'), its position, type, and form are specified.

Table 7. Inspect results of the segment '*because he* belongs to the top players'

Linguistic element	Present?	Type?	Token?	Position?
1. contraction	no			
2. connector	yes	causal	because	initial
3. referent	yes	topic	pronoun	preverb
4. action marker	no			
5. modal marker	no			
6. temporal marker	no			
7. response marker	no			
8. theme marker	no			
9. superstructure marker	no			

Inspection of this segment illustrates that the initial subclause contains a causal connector (*because*) and a referent denoting the topic, which is also in the preverb position (*he*)[31]. Other linguistic elements are not present.

Table 8. Inspect results of the segment '*If the players* are *quarrelling*

Linguistic element	Present?	Type?	Token?	Position?
1. contraction	no			
2. connector	yes	conditional	if	initial
3. referent	yes	local	full NP	preverb
4. action marker	yes	main verb	quarrel	--
5. modal marker	no			
6. temporal marker	yes	verb-inflexion	durative	--
7. response marker	yes	connotation	problem	--
8. theme marker	no			
9. superstructure marker	no			

Note '--': feature does not apply to this linguistic element

Inspection of the segment of fragment (10) shows that this initial subclause contains a conditional connector and an action verb in the gerund which is regarded as a temporal

marker. In addition, the segment contains a response marker indicating a problem (*to quarrel*).

In the same fashion, an entire text can be inspected for the presence of relevant linguistic elements. The inspection results of text 1, on Ed Nijpels, are presented in table 9. The linguistic elements inspected by PISA are italicized. The number following the elements denote which type of element it is; the numbers correspond to those presented in table 6. Thus, '*I* (3)' indicates that the element *I* is a referent (3), '*if* (2)' means that it concerns connectors, etc. For practical reasons, the features (type, token, position) of the elements present are not indicated.

Table 9. INSPECTed linguistic elements of text 1 (written by a young adult, university level)

Segment	linguistic elements present
1	*I* (3) would,
1a	*if* (2) forced to make a choice
1	*like* (5) *to be like* (9) *Ed Nijpels*.
2	*He* (3+9) *looks quite nice* (8),
3	.. (1) is *well dressed* (8),
4	and (2) ..(1) *expresses* (8) *himself* (3) very well.
5	*Besides*, (2) *he* (3) *has undertaken* (4+6) all kinds of activities alongside *his* (3) rather varied *career* (8).
6	It is *this variation* (3) that appeals to *me* (3+9) very much.
7	*His* (3) "main *job*" (8) *at this moment* (6) *is being* (6) *mayor* (8) of Breda.
8	*But* (2) as a *compare on tv* (8) *he* (3) holds his own (3) as well!
9a	*When* (2) *he* (3) *relaxes* (8) *after* (6+8) *his busy work* (8),
9	*he* (3) *can* (5) be found in *the tennis court* (8).
10	*and* (2) *the evenings* (6+8) *he* (3) *likes* (5+8) to be with *friends* (8) or just on *his own* (3) reading a *nice book* (8)!
11	Because of *his* (3) open *character* (8) *he* (3) makes *friends* (3+8) easily
12	*and* (2) .. (1) *makes contacts easily* (8).
13	As *a mayor* (8) *he* (3) tries to stand close to *the people* (3).
14	The *age* (8) of *Ed Nijpels*'s (3)
14a	(1) (round about *40*) (8)
14	is an age which has many advantages:
15	One *has had* (6) the chance (and the time!) to make something of one's life.
16	*One* (3) does not be a 'he-man' *anymore*.
17	*Summarizing*(9), *I* (3+9) *must* (5) say that in any case it looks as if *Ed Nijpels* (3) feels comfortable
18	*or* (2)
18a	(2) to put it informally
18	'*he* (3) likes who *he* (3) is'
19	*and* (2) *thàt* (3) is something worth *striving for* (9).

These INSPECT results, together with the relevant features (type, token, and position in the segment), are passed on to CONNECT and to INTEGRATE. The linguistic elements relevant to the microstructure - contraction, connector, referent, and the four types of markers - are sent to CONNECT. Thus in the case of text 1, table 9, CONNECT receives the information that segments 1, 2, and 4 (among others) comprise a referent to the topic, that segment 8 and 9a (among others) comprise a connector, etc.. The linguistic elements relevant to the macrostructure and superstructure are sent to INTEGRATE. In the case of text 1, this function receives the information that segments 5, 7 and 11 include a theme marker, and segment 1 and 17 (among others) include a superstructure marker.

CONNECT

The CONNECT-component decides on the hierarchical position of a segment in the microstructure and on the associated relational meanings. The results of the inspection are matched to the conditions of the condition-action pairs. These condition-action pairs are applied in a fixed order, which, more or less, corresponds to the order of linguistic elements presented in table 6. If a condition is true, the action consists of a decision on linking the segment to the microstructure and on labelling its relational meaning. Following these decisions, the analysis continues with a search for alternative linkings. These may be needed, for example, in the case of an isolated linking. If main decisions result in an isolated linking, INTEGRATE may decide on an alternative.

The main characteristic of CONNECT is that it does not decide on individually inspected elements, but analyses linguistic elements in their combined form.

Take, as an example, the initial subclauses in the fragments (9) and (10). Although they are both initial and contain connectors, they are treated differently. This is because the subclause in fragment (10) also includes other linguistic elements which influence the final linkage.

The INSPECT-results show that the subclause is causally related to the main clause. In CONNECT, this matches with the condition of the rule presented in (11), which states that '*if* a segment contains a connector of the position initial subclause, and of the causal type, the following action should be executed.'. The action to be executed is 'subordinate current segment to the governing main clause'.[32] The referent present in the segment is overruled by the connector.

(11) *if* (connector <position = initial subclause>
& <type = subordinating>
& <token = causal>)
then subordinate segment to main clause governing the current segment,
and label relation as 'cause'.

In comparing this decision what is decided with regard to fragment (10), which is also an initial subclause with a connector, a quite different result can be seen. This is due to the fact that (10) also consists of quite different linguistic elements which influence the final linkage. In addition to the connector within the initial subclause, this segment also has an action verb denoting a problem (the verb to quarrel indicates a negative situation), a non-topic referent, and a temporal marker (the gerund form of the main verb). Some of these elements make the condition of the rule presented in (12) true. The connector is of the type 'initial subclause' and of the token 'conditional'; in addition, the segment contains a response marker.

(12) *if* (connector <position = initial subclause> & <token = conditional>)
and (response marker = true)
then open response-line and go to referent-rules to find a link
and label relation as 'problem'.

The action to be executed is to open a response-line in the microstructure and to search for a referent to which this line can be linked. In searching a referent, CONNECT maintains the application of other rules.[33]

It goes beyond the scope of this illustration to describe the final decision of linking in detail. The above-mentioned fragments are only used to illustrate that linguistic elements of the same type (connectors of an initial subclause) may lead to quite different actions because they differ in other respects, for instance, their token (causal versus conditional), or because the segment contains other linguistic elements which influence the text structure. The examples also illustrate that CONNECT gives priority to certain linguistic elements above others. In the case of fragment (10), the presence of the temporal marker is overruled by the presence of the response marker. The determination of this priority of linguistic elements and its contribution to the structure is based on research of different text corpora.

An illustration of which inspected elements in text 1 are of influence on the microstructure is presented in table 10. As the results of the decisions on the microstructure and relational labelling have already been presented in figure 3 and table 3, respectively table 10 presents the elements of influence according to the rules in CONNECT. These elements are given in bold print. This table illustrates that most segments are linked on the basis of referential relations to the topic.[34]

Table 10. CONNECT results of text 1 (written by a young adult, university level)

Segment	primary linguistic element for a microstructure link
1	*I* (3) would,
1a	***if*** (2) forced to make a choice
1	***like*** (5) *to be like* (9) *Ed Nijpels* (9).
2	***He*** (3+9) *looks quite nice* (8),
3	**.. (1)** is *well dressed* (8),
4	and (2) **.. (1)** *expresses* (8) *himself* (3) very well.
5	*Besides*, (2) ***he*** (3) *has* undertak*en* (6) all kinds of activities alongside *his* (3) rather varied *career* (8).
6	It is ***this variation*** (3) that appeals to *me* (3+9) very much.
7	***His*** (3) "main *job*" (8) *at this moment* (6) *is being* (6) *mayor* (8) of Breda.
8	***But*** (2) as a *compare on tv* (8) *he* (3) holds his own as well!
9a	***When*** (2) *he* (3) *relaxes* (8) *after* (6+8) *his busy work* (8),
9	***he*** (3) *can* (5) be found in *the tennis court* (8).
10	*and* (2) *the evenings* (6+8) ***he*** (3) *likes* (5+8) to be with *friends* (8) or just on *his own* (3) reading a *nice book* (8)!
11	Because of *his* (3) open *character* (8) ***he*** (3) makes *friends* (3+8) easily
12	*and* (2) **.. (1)** *makes contacts easily* (8).
13	As *a mayor* (8) ***he*** (3) *tries* (5) to stand close to *the people* (3).
14	The *age* (8) of ***Ed Nijpels***'s (3)
14a	**(1)** (round about *40*) (3+8)
14	is an age which has many advantages:
15	One *has had* (6) *the chance* (3) (and the time!) to make something of one's life.
16	***One*** (3) does not be a 'he-man' *anymore* (6).
17	*Summarizing*(9), ***I*** (3+9) *must* (5) say that in any case it looks as if *Ed Nijpels* (3) feels comfortable
18	***or*** (2)
18a	**(2)** to put it informally
18	'*he* (3) likes who *he* (3) is'
19	*and* (2) ***that*** (3) is something worth *striving for* (9).

INTEGRATE[35]

As the analysis is started, a superstructure and a macrostructure with relevant themes are activated. INTEGRATE obtains linguistic elements indicating linkings to the macrostructure and/or the superstructure from INSPECT.[36]

In the case of text 1, INTEGRATE obtains the information that segment 2 contains a theme marker referring to Appearance, segment 7 to Career, and segment 11 to Character.

The INTEGRATE component decides that segment 1 is the introduction to the text on the basis of the superstructure-marker *would like to be like + name of the Topic*, which is a reference to the assignment. Because segment 2 has a referent to the topic in subject position, this

segment is regarded as the starting point of the Description-section. The marker *Summarizing* is a perfect example of an indicator of the start of the conclusion.

The theme markers and the superstructure markers responsible for the macrostructure and superstructure, respectively, are printed in bold in table 11.

Table 11. INTEGRATE results of text 1 (written by a young adult, university level)

Segment	primary linguistic element for a macro- and superstructure link
1	***I*** (3+9) would,
1a	*if* (2) forced to make a choice
1	*like to be like* (9) *Ed Nijpels* (9).
2	***He*** (3+9) ***looks quite nice*** (8),
3	.. (1) is ***well dressed*** (8),
4	and (2) .. (1) ***expresses*** (8) *himself* (3) very well.
5	*Besides*, (2) *he* (3) *has* undertaken (6) all kinds of activities alongside *his* (3) rather varied *career* (8).
6	It is *this variation* (3) that appeals to *me* (3+9) very much.
7	*His* (3) "main ***job***" (8) *at this moment* (6) *is being* (6) ***mayor*** (8) of Breda.
8	*But* (2) as a ***compare on tv*** (8) *he* (3) holds his own as well!
9a	*When* (2) *he* (3) ***relaxes*** (8) ***after*** (6+8) ***his busy work*** (8),
9	*he* (3) *can* (5) be found in ***the tennis court*** (8).
10	*and* (2) ***the evenings*** (6+8) *he* (3) ***likes*** (5+8) to be with ***friends*** (8) or just on *his own* (3) reading a ***nice book*** (8)!
11	Because of *his* (3) open ***character*** (8) *he* (3) makes ***friends*** (3+8) easily
12	*and* (2) .. (1) ***makes contacts easily*** (8).
13	As ***a mayor*** (8) *he* (3) *tries* (5) to stand close to *the people* (3).
14	The ***age*** (8) of *Ed Nijpels*'s (3)
14a	(1) (round about ***40***) (3+8)
14	is *an age* which has many advantages:
15	*One* (3) *has had* (6) *the chance* (3) (and the time!) to make something of one's life.
16	*One* does not be a 'he-man' *anymore* (6).
17	***Summarizing***(9), ***I*** (3+9) *must* (5) say that in any case it looks as if *Ed Nijpels* (3) feels comfortable
18	*or* (2)
18a	(2) to put it informally
18	'*he* (3) likes who *he* (3) is'
19	*and* (2) *thàt* (3) is something worth ***striving for*** (9).

4.6 Evaluation of PISA

The basic principles for PISA were described in Sanders 1992 and Sanders and Van Wijk 1995a. It was described as a procedural approach to analyze text structure on the basis of an extended set of cohesive linguistic elements which yielded a hierarchical structure (microstructure) and a labelling of the relations.

The further development of PISA has left these principles unaffected. The improvements consist of *extensions* and a *reorganization* of the functions used in assigning the structure (including different labellings).

Whereas the first release focused on developing a microstructure, in this release, the necessity of a macro- and superstructure level are taken into account as well.[37] These levels are especially important in psycholinguistic research, as text structure at all levels provides insight into conceptual operations and representations. This extension on the output side implied, however, that more world and discourse knowledge were needed on the input side. As a consequence, the possibilities for implementation became restricted, as it is rather difficult to formulate strict procedural rules for the assignment of a macro- and a superstructure (as more knowledge has to be taken into account).

The architecture of PISA was reorganized, in order to create a more flexible structure. This new architecture made it possible to extend or adapt the set of functions, which is obviously of great importance in the further development of PISA. Application to other text types, for instance, requires new rules, in particular, for the macro- and superstructure (see, for an application to judicial texts Schilperoord 1995 and Schilperoord & Van der Pool 1995).

Notes

1. In many studies on language development, the length of the text and the sentence structure (e.g. Hunt 1965) are used as indicators for developmental changes. In developmental research this is not an adequate operationalization, because the length of a text and the sentence structure do not reflect the same conceptual processes. Therefore in this study text structure was analyzed.

2. Meyer distinguishes three primary levels of expository text (1985: 17). The first is a microproposition level, which is concerned with the relations between and within individual sentences. The second is the macropropositional level, which concerns the relationships among ideas represented in complexes of propositions or paragraphs. The third level is the top level, the level corresponding to the overall organizing principle. For this term we use Van Dijk's term 'superstructure' (Van Dijk 1986).

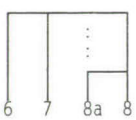
3. This does not mean that we also want to include implicit relations. The aim of the analysis is to deliver a structure on the basis of 'shallow' processing (see also chapter 2). Explicit markers are relied upon as much as possible.

4. Meyer (1985: 17) calls these patterns *collection*.

5. For a theory on temporal markers and their references points see, for instance, Oversteegen 1990.

6. They state that if-clauses preceding the main clause may open Response lines. These Response lines consist of a goal or a problem which has to be reached, or solved.

7. In addition, this segment is an event in the narrative pattern and, therefore, has a secondary link to the line indicating the sequence. This secondary link is graphically represented by a dotted line as is shown below (for conventions, see Sanders & Van Wijk 1995b).



8. An RST analysis is made up of schemas, which indicate how a particular text unit is decomposed into other units. Each schema consists of a *nucleus*, and a *satellite* which is linked to the nucleus by a certain relational meaning (Mann & Thompson 1987; 1988; for a discussion on nuclearity, see 1988: 265-269).

9. It is compatible to the taxonomy developed by Sanders, Spooren and Noordman 1992; see also Sanders (1992: chapter 6).

10. The two basic primitives within LDM for building the discourse structure are *discourse constituent units* (dcu's), consisting of clauses and carrying the propositional content, and *discourse operators*, which modify one or more dcu's by providing information on the state of the discourse and the relation of discourse entities to one another without carrying propositional information themselves. There are three types of discourse operators: assigners, connectors, and push/pop markers (see, for further details, Polanyi 1988: 605-606).
The context for the dcu's consists of higher level structures, called *discourse 'genre' units* (DU's). These DU's and dcu's are the basis for constructing *Speech events*, which in turn, constitute *Interaction*. The LDM is based primarily on oral discourse (doctor/patient interactions). This is also expressed in the four types of dcu's which are distinguished: the sequence, the expansion unit, the binary structure, and the interruption.

11. The possibilities for linking a dcu to the discourse structure in LDM is restricted to the most righthand suitable accessible node in the structure in an existing discourse parse tree. This means that returning to previous dcu's is only possible if pop-markers are used (see Redeker 1991).

12. These discourse schemes are equivalent to what McKeown 1985, working from a computational approach, calls 'schemata'. According to McKeown a schema "defines a particular organizing principle for text and is used to structure the information ... It is used to guide the generation process, controlling decisions about what to say when in a text" (1985: 19-20).

13. According to Meyer's distinction of levels of prose, discourse schemes are operationalizations on the macro level, and the superstructure expresses the top-level structure (Meyer 1985).

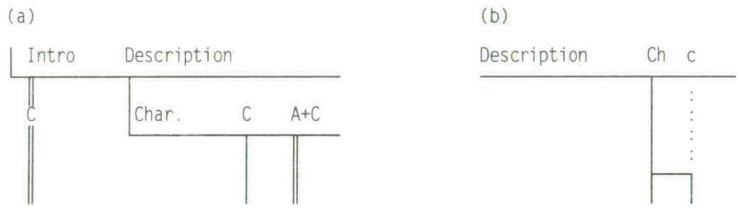
14. The relational meanings that PISA assigns are based on the taxonomy of Sanders, Spooren & Noordman 1992.

15. This is crucial in the interpretation of the incrementally built structure, see chapter 5. This presentation is quite different from the way in which Polanyi 1988 represents structures. In contrast to our notation, the nodes in the LDM parser do not always correspond to text segments and their order may be neglected as well.

16. See Van Dijk 1986, for a partitioning of news reports; Kintsch & Van Dijk 1978 for a partitioning of research reports; Toulmin 1958, for a partitioning of argumentative texts; Van Wijk & Van der Pool 1990, for a partitioning of business letters.

17. These themes correspond to the themes mentioned in the assignment presented to the subjects.

18. Most segments are linked to a single theme, as is the case in both examples. A segment may, however, be linked to two (or more) themes. This is the case if that segment is thematically related to two different themes. Such a multiple linking is represented in the macrostructure by a double line; both themes are indicated on the Description line, in the order in which they occur in the segment. If a theme is mentioned in the introduction of a topic, this is denoted by a double line in that case; the theme-abbreviation is added to the Introduction segment likewise. See (a) for an example of these multiple linkings.
In addition, a subclausal segment may be subordinated in the microstructure, whereas it relates to a theme differing from the one in the main clause. In that case, the relation to the macrostructure is represented by a weak linking, which is indicated by a dotted-line; the theme is indicated on the Description line by a lower case abbreviation, see (b).

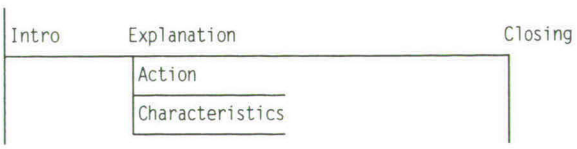


19. The meanings of the relations influence the building of the structure. They are not simply an artefact of the hierarchical aspects.

20. PISA hinges on existing parsers.

21. This was also pointed out in the example of the rocket being launched in section 3.3.1.1.

22. The superstructure varies over writing tasks in the sense that the central part has a different function and may, therefore, have other lines (see Sanders & Van Wijk 1995a,b for an example of the Action line and Characteristic line in explanatory texts). The superstructure for explanatory texts can be depicted as follows:



23. It is deduced from world knowledge in the form of scripts (Schank & Abelson 1977, Bower, Black & Turner 1979). However, the use of world knowledge is restricted to the encyclopedic type, as such script-like knowledge is much too complicated to be made explicit.
This knowledge used in the analysis of descriptions of persons has been inductively formulated on the basis of the corpus used. It consists of a list of items indicating a theme, which can be successfully applied in most of the cases.

24. The reading process described in chapter 3 formed a starting point for formulating the requirements on which PISA is based. It should be stressed that no claims are made concerning the cognitive plausibility of the operations of PISA.

25. See section 4.2.1 in which it is argued that Halliday and Hasan's inventory needs to be extended by adding the concept of global structure (a).

26. See section 4.2.1 in which it is argued that Halliday and Hasan's inventory needs to be extended in order to take into account more divergent aspects within the domain of cohesive features (c).

27. This requires a recognizable relation between words. Halliday and Hasan distinguish between reiteration (including identity of reference) repetition, superordinates, subordinates and synonyms, and collocation, the semantic relationships between words that often co-occur.

28. See also section 4.2.1 in which it is argued that Halliday and Hasan's inventory needs to be extended by enlarging the domain with other cohesive features (b).

29. This method of deriving a meaning is similar to inferencing; it differs with that technique that in the cases described here, an explicit expression is present and needs not be inferred.

30. In addition, it should be noted that different formulations with respect to the certainty of the information presented are treated equally. Segments such as (a) are regarded in the same fashion as those like (b). In addition to these variants, formulations such as (c) also occur.

    (a) I think he is 20 years old.
    (b) He is 20 years old, I think.
    (c) He is about 20 years old.

    These differences in formulations are assumed to be the result of Shape operations, and are, therefore, beyond the scope of this study.

31. Actually, INSPECT also examines the position of the antecedent of the referent. This referent has its antecedent in the segment preceding the fragment in the preverb position. As it is not relevant in this illustration it is left out of consideration in the example.

32. The segments presented in (9) and (10) are taken care of by rule 'Connector-7' (see Appendix to chapter 4). To illustrate the general idea of linking segments, this rule has been presented in an elaborated form, as in (11) and (12).

33. In order to link the segment, rule (12) refers to another rule which searches for a referent. As this subclause does not contain a referent with an antecedent in the text, a referent has to be found in the main clause. On the basis of that referent, the initial subclause is linked to the structure. The information about the response marker is used to reveal that this is not an ordinary initial subclause, but one which initiates a response pattern.

34. Strictly speaking, most linguistic elements inspected for and tested by CONNECT are not present in the corpus of this study. PISA, however, was developed with the aim of analyzing expository texts, including descriptive ones (Mosenthal 1985). As this chapter describes PISA in its general form, it includes several inspections and CONNECT rules which do not apply in the present study.

35. It is possible to create a text structure solely on the basis of INSPECT and CONNECT. This results in a microstructure with labelled relations. A more sophisticated (and psychologically more plausible) result is obtained if integration takes place with world and discourse knowledge.
An option for a more sophisticated architecture of PISA is to make the INTEGRATE component '*interactive*' and order INSPECT to look for additional triggers, if certain triggers of a macrostructure and/or a superstructure are found by INSPECT.

36. INTEGRATE still has a 'static' method of processing, due to the incremental way in which PISA operates. One of the aspects for further development would be to convert this static method into a dynamic method of processing, which looks back and ahead when assigning a structure. In that case, INTEGRATE would have to ask INSPECT to examine whether certain features were present if a certain superstructure were to be activated.

37. In addition, the Response lines are developed in greater detail (see Sanders & Van Wijk 1995a,b).

5 TRACE: a reconstruction of conceptual activities

5.1 Introduction

In terms of empirical research, the structures PISA yields are raw data. They need to be reduced and modelled in order to provide insight into the conceptual processes in writing a text. As writing processes leave traces in the text, the PISA structures may be regarded as prints of (final) conceptual activities (see chapter 3, section 3.3.2). On the basis of this assumption, the PISA structures were used to reconstruct parts of Reflect and Select operations.[1] The reconstruction resulted in a TRACE model, in which TRACE is an acronym for: Text-based Reconstructions of Activities in the ConceptualizEr. The TRACE model for the corpus used in this study will be presented in this chapter.

This step from analyzing to modelling is the central issue in this chapter (see the marked box in figure 1).

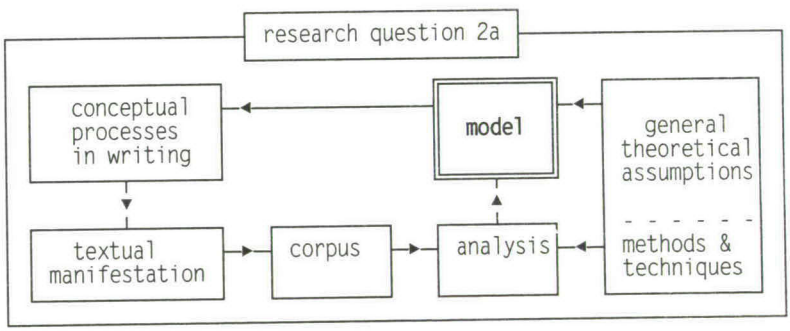


Figure 1. The position of this chapter in the study

A TRACE model represents conceptual operations in terms of a set of reconstruction rules. Each model consists of two main rules, irrespective of the type of text. One rule corresponds to Reflect operations, and the other to Select operations. The actual specifications of these main rules vary for different types and topics of texts. The architecture of the system, two main rules governing a set of embedded rules, remains unaffected.[2]

In this chapter, a TRACE model for descriptive texts is presented. Before the reconstruction

rules are presented in section 5.3, the theoretical assumptions which underlie the reconstruction are discussed in section 5.2. The latter section also briefly describes how the information needed to make the reconstructions is induced from the PISA structures. The scope and the restrictions of the rules are evaluated in section 5.4.

5.2 Backgrounds of the TRACE model

5.2.1 General theoretical assumptions

A TRACE model is based on several theoretical assumptions. As was discussed in the theoretical framework in chapter 2, the conceptual part of writing consists of an elementary Select operation responsible for the knowledge-telling way of text production (Bereiter & Scardamalia 1987).[3] This Select operation consists of suboperations, namely, the retrieval and organization of content, of which retrieval is the most prominent. The organization of information is a residual product of using retrieval cues in a certain order. Writing is reduced to a routine, primarily concerned with what to say next and how to put it into appropriate language. It mainly concerns local considerations and does not involve operations such as goal setting (Bereiter & Scardamalia 1987: 300).

Select may be governed by Reflect, a problem-solving activity, which guides and controls this basic operation (cf. Bereiter & Scardamalia's knowledge-transforming model 1987, discussed in chapter 2, section 2.2.2). Reflect considers content related problems as well as rhetorical ones.[4] It results in focusing on ideas in separate information units, as a result of dealing with *content problems*. It also yields a global organization of the message by attending more closely to the order in which the information will be mentioned (Bereiter & Scardamalia 1987: 314).[5] Dealing with such *rhetorical problems* forms the strategic aspect of Reflect; it is strictly bound to (written) communication (o.c. 1987: 302).

This relation between Select and Reflect can be represented in a more formalized fashion (see (1)).

(1) (Reflect (Select)) → Message

Select operations are embedded in Reflect operations. Together, they result in the *Message* the writer wants to convey.[6]

The activities that take place in Select yield a string of information-units that have the form of arguments, characteristics, events, etc. Each of these units can be elaborated. In a coherent text the information units are related to each other by coherence relations such as addition, sequence, and contrast.

By guiding and controlling Select on the basis of intentions and expectations, the Reflect operator has two consequences which leave observable traces on different levels in the final text. Firstly, the message may contain evaluative statements concerning specific information; this results from solving content problems ('What do I mean?'). Secondly, the message may be framed by introducing and concluding the information retrieved by Select; this results from solving rhetorical problems ('How do I say what I want to say?').

Levelt raises a relevant question in this regard (1989: 124): "If macroplanning (that means Select in the terminology of this study) is procedural in nature, what form do these procedures have?".[7] This question has been central in studies within the research field of artificial intelligence (see, for instance, Appelt 1985) and computational linguistics (see, for instance, Paris 1993), and has resulted in models of sentence planning and text generation, respectively.

This study is aimed at developing a psycholinguistic model of the conceptual processes of individual writers.[8] It is therefore desired, that Reflect operations as well as Select operations be described, although the analysis of communicative intentions (part of the Reflect operator) is complicated greatly by the fact that there are many mappings from intentions to 'speech acts' (Levelt 1989: 123).

Given the above-mentioned theoretical assumptions, PISA structures are an important data source for inferring how writers may have operated. As they are regarded as prints of conceptual activities in a comparable format, patterns in the structures can be used to induce regularities in Reflect and Select operations. The claims that underlie the interpretation for *descriptive* texts are summarized in table 1.[9]

Table 1. Reflect and Select operations and their manifestations in the PISA structure of descriptive texts[10]

Conceptual operation	Manifestation in the PISA structures of descriptive texts
Select	- themes in the macrostructure represented on the characteristics line - microstructure - second mentionings of the themes in the macrostructure (discontinuities)
Reflect	- parts within the superstructure (such as introduction, conclusion) - explicitly stated comments about the information presented (marked with a 'c')

It is assumed that operations within *Select* are retraceable in PISA's macro- and micro structure, as well as in the relational meanings. The macrostructure of PISA mirrors the (different) themes from which the information has been retrieved. The size of the themes (number of segments per theme), and the depth of the structure (displayed by the hierarchical layers within the microstructure) show the quantity of the information retrieved. The presence or absence of discontinuities in the description of themes represent the order in which the information has been retrieved.[11]

It is assumed that *Reflect* is manifested in a PISA structure by an introduction and a conclusion in the superstructure, and by qualifications of the person described, indicating an evaluation or a metastatement.

Deriving a process description from these structural features is an iterative and interpretative research step. By comparing PISA structures and searching for similar patterns, regularities have to be accounted for in terms of underlying rules. Every node in the PISA structure is regarded as (part of) a conceptual activity, a decision with regard to the manner in which the text production is continued.[12] All these activities should be accounted for in TRACE.

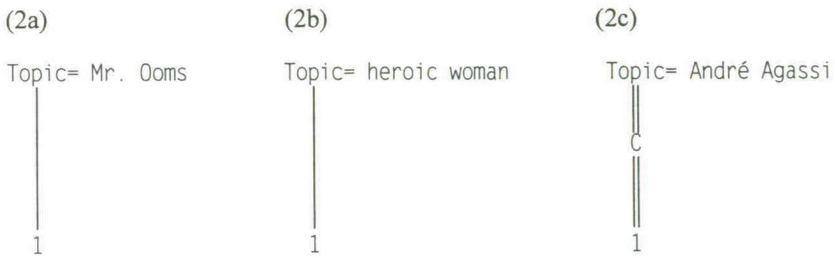
The format chosen for specifying the conceptual activities is a set of *decision rules*. The idea behind these rules is that text structures consist of stereotypical parts. The rules make the hierarchical relations between these parts explicit.

5.2.2 The induction of the TRACE model for descriptive texts

The empirical basis for developing the TRACE model consisted of the PISA structures of the texts in the corpus. The structures were thoroughly examined, irrespective of the topics of the texts (concrete or constructed), the age of the subjects, and the type of school. Each structure was scrutinized to detect similarities with other structures. On the basis of similar patterns, Reflect and Select decisions, presumably made during text production, were reconstructed. Two examples of inducing decision rules are given to illustrate this induction.

Example 1

In (2a to 2c) three different beginnings of PISA structures are presented, which display the introduction of the topic.



The topic indications within these three partial structures show that there are two ways in which the topic has been introduced in the first segment of the text:

- by mentioning the proper name of the protagonist (*Mr. Ooms* in (2a), *André Agassi* in (2c)), or
- by stating a stereotype, which is an anonymous instantiation of a type of person with prominent characteristics (*heroic woman* in (2b)).

This type of introduction may be supplemented with information related to a specific theme, as is the case in (2c). The double line in (2c) marked with a 'c' indicates that the writer has added a salient feature related to the theme Career (*the tennis player André Agassi*).

Nowhere in the corpus did a writer introduce a topic by mentioning a stereotype as well as a name (the stereotype was defined as anonymous). That is why introducing the topic leaves open two possibilities (see (3)).

(3) INTRODUCE Topic → Mention name or Stereotype

As (2c) shows, a writer can also add a salient characteristic to the topic mentioned. Thus, (3) should be completed as has been done in (4).

(4) INTRODUCE Topic → Mention name or Stereotype
+ [Saliency]

Application of this rule to the examples (2a to 2c) yields the reconstructions presented in table 2.

Table 2. Reconstruction of conceptual activities in writing the examples (2a to 2c) by applying the TRACE rules presented in (4)

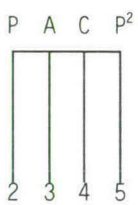
Example	Result of application
(2a)	INTRODUCE Topic → Mention name
(2b)	INTRODUCE Topic → Stereotype
(2c)	INTRODUCE Topic → Mention name + Saliency

Example 2

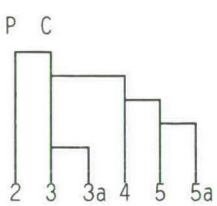
The examination of the PISA structures of the whole corpus in close connection with the content yielded an enumeration of aspects described per theme. A description of the Character theme consisted of aspects related to the topic's behaviour or capacities. Segments describing the Appearance presented aspects of the topic's face or body; a discussion of the Career theme consisted of aspects of the topic's education, profession, or his ambitions; addressing the topic's Hobbies consisted of a description of the topic's sports, creative activities or collections. The Personal particulars theme presented the topic's personal particulars, idiosyncrasies, talents, and properties.[13]

In (5a) and (5b) two fragments are presented of the Description-section in two different PISA structures.[14]

(5a)



(5b)



Structure fragment (5a) represents a list of characteristics related to different themes, those being Personal particulars, Appearance, Career, and again Personal particulars, respectively.[15] The descriptive part represented in (5b) presents two characteristics belonging to different themes, Personal particulars and Career, of which the latter has been elaborated by segments 3a to 5a.

These structure fragments show that the writers have used different ways of describing a

topic. The simplest of these is (5a). The writer characterizes the topic by presenting theme-related characteristics, one after the other. The characterization of the topic in (5a) can be reconstructed by the rules represented in (6).

(6)
a.	CHARACTERIZE Topic	→ CHOOSE THEME
b.	CHOOSE THEME	→ Character, Appearance, Career, Hobbies, Personal Particulars
c.	Character	→ Present an aspect of the topic's Behaviour, Capacities, ...
d.	Appearance	→ Present an aspect of the topic's Face, Body, ...
e.	Career	→ Present an aspect of the topic's Education, Profession, Ambitions, ...
f.	Hobbies	→ Present an aspect of the topic's Sports, Creative activities, Collections, ...
g.	Personal particulars	→ Present the topic's Personal particulars, Idiosyncrasies, Talents, Properties, ...

Fragment (5b) shows that the writer has elaborated one characteristic several times ((S3) *he has his own band:* (S3a) *'The Jumping Jewels'.* (S4) *It is a superb band* (S5) *It consists of 5 persons* (S5a) *one bass guitarist, 1 supervisor and 2 ordinary guitarists and in addition a drummer.*) Thus, rule (6a) which reconstructs the characterization of the topic should be supplemented by an optional function which deals with any possible elaborations (see (7)). Each aspect presented may be elaborated on, for instance, through the use of a specification or a reason. As this may occur more than once, the rule for elaborating on a theme is recursive (this is marked with an asterisk; the optional use of the function is represented by [...].)

(7)
a.	CHARACTERIZE Topic	→ CHOOSE THEME + [ELABORATE THEME]*
b.	CHOOSE THEME	→ Character, Appearance, Career, Hobbies, Personal Particulars
c.	Character	→ Present an aspect of the topic's Behaviour, or Capacities, ...
d.	Appearance	→ Present an aspect of the topic's Face, Body, ...
e.	Career	→ Present an aspect of the topic's Education, Profession, Ambitions, ...
f.	Hobbies	→ Present an aspect of the topic's Sports, Creative activities, Collections, ...
g.	Personal particulars	→ Present the topic's Personal particulars, Idiosyncrasies, Talents, Properties, ...
h.	ELABORATE THEME	→ Specify, give a Cause, ...

In this inductive fashion decision rules specifying (final) Reflect and Select operations were derived. The complete set of reconstructed rules is presented in the next section.

5.3 A TRACE model for descriptive texts

In the case of the production of a descriptive text, the SELECT function represents procedures for describing a topic (DESCRIBE Topic), the main function REFLECT consists of the evaluation of that description (EVALUATE (DESCRIBE (Topic))).[16] These main procedures of the TRACE model for descriptive texts are presented in table 3.

Table 3. The main procedures of the TRACE model of descriptive texts

REFLECT (SELECT)	→	EVALUATE (DESCRIBE Topic)
DESCRIBE Topic	→	INTRODUCE Topic
	+	CHARACTERIZE Topic*
EVALUATE (DESCRIBE Topic)	→	[EVALUATE (INTRODUCE Topic)]*
	+	EVALUATE (CHARACTERIZE Topic)*
	+	[EVALUATE (Topic and/or Description)]*

Key:
- CAPITALS : a function with its argument
- * : a recursive use of a function
- [ ] : an optional use of the function
- + : juxtaposition of functions

A description starts with the introduction of the topic (INTRODUCE Topic), followed by a characterization of the topic (CHARACTERIZE Topic), which consists of several characteristics.

This DESCRIBE function may constantly be directed by the EVALUATION function. This evaluative function takes into consideration the introduction of the topic (EVALUATE (INTRODUCE Topic), as well as the characterization of the topic (EVALUATE (CHARACTERIZE Topic). In addition, this function also evaluates the topic and the final description generated (EVALUATE (Topic and/or Description)).

Each of these functions consists of subfunctions. The subfunctions of DESCRIBE Topic are specified in table 4, those of EVALUATE (DESCRIBE Topic) in table 5. Brief explanations of each are given below.

Table 4. A specification of DESCRIBE Topic

DESCRIBE Topic	→ INTRODUCE Topic + CHARACTERIZE Topic*
INTRODUCE Topic	→ Mention name or Stereotype + [Saliency]*
Mention name	→ Present proper name of the protagonist
Stereotype	→ Give a one-word stereotype of the protagonist
Saliency	→ Mention a distinctive characteristic related to a theme
CHARACTERIZE Topic	→ [ADD CONTEXT] + CHOOSE THEME + [ELABORATE THEME]
ADD CONTEXT	→ Present background or Condition to characteristic or Mark next theme explicitly
CHOOSE THEME	→ Character or Appearance or Career or Hobbies or Personal particulars
Character	→ Present an aspect of the topic's Behaviour, Capacities, ...
Appearance	→ Present an aspect of the topic's Face, Body, ...
Career	→ Present an aspect of the topic's Education, Profession, Ambitions, ...
Hobbies	→ Present an aspect of the topic's Sports, Creative activities, Collections, ...
Personal particulars	→ Present the topic's Personal particulars, Idiosyncrasies, Talents, Properties, ...
ELABORATE THEME	→ Specify, give a Cause, present an Event, ...

In describing a person, a writer starts by introducing the topic he has chosen (INTRODUCE Topic).[17] Introducing the topic consists in any case of mentioning the protagonist's proper name or stereotyping the person in one or two words. Textual manifestations on which these rules are based are presented in (8) and (9).[18] They are often formulated in a phrase referring

to the assignment '*I would like to be like ...*'. A writer may extend this indication of the topic by adding one or more salient characteristics, the protagonist's career for instance, as in (10).

(8) I would like to be like Mr. Ooms.
Ik zou graag willen lijken op meneer Ooms.

(9) I would like to be like a heroic woman.
Ik zou willen gelijken op een heldhaftige vrouw.

(10) I would like to be like the singer Adamo.
Ik zou graag willen gelijken op de zanger Adamo.

The mentioning of a topic is an introduction to characterize this topic (CHARACTERIZE Topic). In characterizing a topic, a writer must choose a theme to describe this topic (CHOOSE THEME). A discussion of a theme may be related to one of the following five types: Character, Appearance, Career, Hobbies, or Personal particulars.

Describing a person's *Character* consists of information about his behaviour (is the person brave, calm, etc.) or capacities (is the person intelligent, technically inclined, etc.). An example of a textual result is presented in (11).

(11) He has a friendly character.
Hij heeft een sympathiek karakter.

The theme *Appearance* includes a description of the protagonist's face (colour of the eyes, form of the mouth, etc.) or body (his length, the kind of clothes he wears, etc.). A textual manifestation is presented in (12).

(12) He has dark hair.
Hij heeft donker haar.

A discussion of the theme *Career* provides information about the person's education (primary, secondary etc.), profession (function or activities), or ambitions (related to a profession, for instance, the wish to become a famous actor). In (13) an example of the final result is presented.

(13) His profession is police-officer.
Hij is van beroep politie-agent.

Discussing the theme *Hobbies* includes the sports an individual likes to participate in (for instance, swimming, playing badminton), activities of a creative nature (making music or drawing) and collecting objects (stamps, sugar bags, etc.). An example of the textual manifestation is given in (14).

(14) His hobby is American football.
Zijn hobby is Amerikaans voetbal.

The information about *Personal particulars* consists of presenting data such as place and date of birth, idiosyncrasies (nail-biting), talents (being the best in class), and possessions (owning a house, a telescope or an American kitchen). An example of the textual result is presented in (15).

(15) He is 40 years old.
Hij is 40 jaar oud.

Each theme discussed may be extended by the operations (ADD CONTEXT) and (ELABORATE THEME).[19] The addition of contextual information usually precedes the discussion of a theme, though it does not necessarily have to do this. A contextualization may be constructed by presenting background information or a condition to a characteristic (see (16a)). A writer may also mark the next theme explicitly by an introductory statement of what will be discussed next (see (17a)).[20]

(16) a. *If he has spare time*
Als hij vrij heeft

b. he likes playing football.
speelt hij graag voetbal.

(17) a. *Now for something about his appearance.*
Nu iets over zijn uiterlijk

b. I think he is very good looking.
Ik vind hem erg knap.

A writer may extend a characteristic by one or several elaborations (ELABORATE). He may decide to specify a previously mentioned characteristic, to substantiate a statement with a

reason or to present an event for illustration.[21] Textual examples of each of these elaborations are presented in respectively (18), (19), and (20).

(18) a. He has his own band: 'The Jumping Jewels'.
Hij heeft een eigen orkest: 'The jumping Jewels'.

b. *It is a superb band.*
Het is een geweldig orkest.

(19) a. He does not have an ugly face
Hij heeft geen lelijk gezicht

b. *for otherwise the girls would not look at him.*
want anders zouden de meisjes niet naar hem kijken.

(20) a. They play all over the world
Ze spelen over de hele wereld

b. *and they played in the Netherlands once.*
ze hebben ook eens in Nederland gespeeld.

The DESCRIBE function may be evaluated by EVALUATE (DESCRIBE Topic). The subfunctions of the operation governing DESCRIBE are presented in table 5. A brief explanation is also given.

EVALUATE affects the ongoing execution of DESCRIBE Topic in several ways. The evaluative function is applied in a predetermined way to the two subfunctions of DESCRIBE Topic, i.e., (EVALUATE (INTRODUCE Topic) and EVALUATE (CHARACTERIZE Topic)), and adds a separate evaluation to the results of these subfunctions, which are referred to as *Topic* and *Description.*

Table 5. A specification of EVALUATE (DESCRIBE Topic)

EVALUATE (DESCRIBE Topic)	→ [EVALUATE (INTRODUCE Topic)]* + EVALUATE (CHARACTERIZE Topic)* + [EVALUATE (Topic and/or Description)*]
EVALUATE (INTRODUCE Topic)	→ INTRODUCE (Topic) + [EVALUATE (INTRODUCE)] + [EVALUATE (Topic)]
EVALUATE (INTRODUCE)	→ Show positive or negative attitude towards assignment
EVALUATE (Topic)	→ Give positive or negative argument for topic choice; Show personal attitude towards the topic; Summarize main characteristics
EVALUATE (CHARACTERIZE Topic)	→ CHARACTERIZE Topic + [GENERAL-EVALUATION (CHARACTERIZE Topic)] + [SPECIFIC-EVALUATION (CHARACTERIZE Topic)]
GENERAL-EVALUATION	→ Give opinion about entire topic, either positive or negative, during the execution of (CHARACTERIZE Topic); Make a general meta-textual remark during the execution of (CHARACTERIZE Topic)
SPECIFIC-EVALUATION	→ Give opinion about specific characteristic of topic from writer's or reader's perspective during the execution of (CHARACTERIZE Topic); Make a specific meta-textual remark during the execution of (CHARACTERIZE Topic)
EVALUATE (Topic and/or Description])	→ [EVALUATE (Topic)] + [EVALUATE (Description)]
EVALUATE (Description)	→ Conclude description with metastatement

The introduction of the topic may be evaluated in two ways. First, the fact that the writer is forced by the assignment to introduce someone may be evaluated (EVALUATE (INTRODUCE)), and second, the topic which has been chosen may be given consideration (EVALUATE (Topic)). The latter consists of presenting positive or negative arguments for the topic. These may precede the characterization or may follow it. If the evaluation of the topic follows the characterization it may consist of showing a personal attitude towards the topic, or of summarizing the main characteristics. Examples of textual results of these reflections are presented in respectively (21), and (22).

(21) I would like to be like Linda de Mol, *because she always seems so friendly, and because she is very successful.*
Ik zou graag willen lijken op Linda de Mol omdat ze altijd zo vriendelijk overkomt en omdat ze veel succes heeft.

(22) I would like to be like Willem van Hanegem. *Though not so much because of his appearance, as he is not very handsome. Besides, I do not like his curls either. The reason why I would like to be like him is that he is good at just being himself.*
Ik zou graag willen lijken op Willem van Hanegem. Niet zozeer vanwege zijn uiterlijk, want zo knap is hij nu ook weer niet. Bovendien vind ik die bos krullen ook niet aantrekkelijk. De reden waarom ik wel op Van Hanegem zou willen lijken, is dat Van Hanegem zo heerlijk zichzelf kan zijn.

Considerations about why one should actually mention a topic (EVALUATE (INTRODUCE)) may lead to a positive or negative attitude towards the assignment. Examples of phrases expressing these kinds of evaluations are presented in (23) and (24).

(23) I would, *if forced to make a choice, like to resemble Ed Nijpels.*
Ik zou, als het dan toch moet, wel op Ed Nijpels willen lijken.

(24) *I am rather pleased with myself. Therefore, I do not feel the need to resemble someone else. But in order to fulfil the assignment*, I would like to resemble Candy Dulpher.
Ik ben best tevreden met mezelf. Ik heb daarom niet echt de behoefte om op iemand anders te lijken. Om toch te voldoen aan de opdracht zou ik graag Candy Dulpher zijn.

The evaluation of the characterization of the topic (EVALUATE (CHARACTERIZE Topic)) leads to the addition of general evaluations (GENERAL-EVALUATION) or specific evaluations (SPECIFIC-EVALUATION). The general evaluations are mainly positive, as otherwise the writer would not have chosen this topic. A textual example resulting from these considerations is presented in (25). On the other hand a writer may decide to add his personal opinion to a description of a specific characteristic. He may do this from his own perspective, giving his opinion, or from a reader's perspective by announcing that reader. Examples of phrases expressing these thoughts are presented in (26) and (27).

(25) I get along with him quite well.
Ik kan goed met hem opschieten.

(26) a. He plays football,
Hij doet het liefste voetballen

b. *whereas I prefer to ride horses.*
maar ikzelf doe liever paardrijden.

(27) a. Do not think he has only got one car
nu moet u niet denken dat hij maar een wagen heeft

b. in fact, he has 12 cadillacs and a number of sport and luxury cars.
hij heeft namelijk 12 cadillacs en nog enkele sport- en luxe wagens.

c. *Well, now you know that as well.*
Zo dat weet u ook weer.

The last part of evaluating a description consists of a backward reflection on the topic chosen (EVALUATE (Topic)) and the description presented (EVALUATE (Description)). These considerations should round off the depiction of the person and complete the Message (see (1)). A final evaluation of the Topic consists of presenting an argument - of course, only if this has been not presented previously - or of showing the personal attitude towards the topic, or it may have the form of a summary of the most important characteristics. Examples of these thoughts which may conclude the description are presented in (28) and (29). An evaluation of the entire previous description may consist of a metastatement explicitly stating that the description is finished (see (30)).

(28) *I hope he will and can be playing for a long time*, because I like the Beatles' music.
Ik hoop dat hij nog lang kan en zal spelen, want ik luister graag naar Beatle muziek.

(29) *In conclusion, he is a personality, who has many friends.*
Samengevat, hij is een schone persoonlijkheid die veel vrienden heeft.

(30) *There is not much more that I can tell you about him.*
Meer kan ik niet over hem vertellen.

The rules presented above are used to retrace aspects of Reflect and Select operations. The way these rules are applied to an entire text and its structure is illustrated in the next section.

5.4 An illustration of the TRACE model

In contrast to PISA, which is an instrument for describing *texts*, a TRACE model is used to explore and describe aspects of *conceptual processes* underlying the text structures. The TRACE model transforms the PISA structures into a description of which information has been Selected, in what order and what has been given further consideration. Applying the rules to a text and its structure yield a reconstruction of aspects of Reflect and Select. Such a reconstruction is illustrated by the text presented in table 6 (the Dutch version of this text is presented in the Appendix to chapter 5). The PISA structure is presented in figure 2.

Table 6. Text written by an adult (university level)

1	I would like to be like the tennis player André Agassi.
2	He is about the same age as I am
2a	(25 years old)
3	and earns his money with my hobby,
3a	which is sports.
4	So he is a professional athlete
5	and
5a	because he is one of the top players,
5	he makes a more than decent living.
6	He distinguishes himself in the tennis world by his striking and eccentric appearance and by his skills at the game of course
7a	Despite his sometimes aggressive tactics on court
7	he appears to be very kind-hearted and friendly
8	and that is a trait I appreciate
9	I do not think he has a lot of spare time left
10	but
10a	when he has spare time
10	he surrounds himself with all kinds of luxury
11	and he spends his hard-earned money like water.
12	A nice guy
12a	who wants to work very hard
12b	and still enjoy himself
12c	and sometimes receive the benefits of his labour.

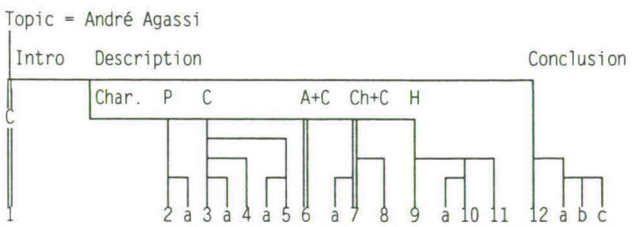


Figure 2. PISA analysis of the text presented in table 6

Applying the rules of the TRACE model for descriptive texts yields the result presented in table 7; the final terms are the segments of the text. (The rules marked with an asterisk are recursive. In order to avoid redundancy, the notation in the table is a shorthand notation of the rules applied; this means not all main rules are repeated.)

Table 7. Reconstruction of conceptual activities in writing the text presented in table 6, on the basis of the PISA structure presented in figure 2. The output denotes the segments of the text

TRACE rule			Segment
EVALUATE (DESCRIBE Topic)			
→ **INTRODUCE Topic**			
+ **EVALUATE (CHARACTERIZE Topic)***			
+ **EVALUATE Topic***			
INTRODUCE Topic			
→ Mention name + Saliency (Career)			1
EVALUATE (CHARACTERIZE Topic)			
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→	Personal particulars	2
→ ELABORATE THEME	→	Specify 2	2a
→ CHOOSE THEME	→	Career	3
→ ELABORATE THEME	→	Specify 3	3a
→ ELABORATE THEME	→	Specify 3	4
→ ADD CONTEXT	→	Background to 5	5a
→ ELABORATE THEME	→	Specify 3	5
→ CHOOSE THEME	→	Appearance + Career	6
→ ADD CONTEXT	→	Background to 7	7a
→ CHOOSE THEME	→	Character + Career	7
→ SPECIFIC-EVALUATION	→	Positive opinion	8
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→	Hobby	9
→ ADD CONTEXT	→	Condition to 10	10a
→ ELABORATE THEME	→	Specify 9	10
→ ELABORATE THEME	→	Specify 9	11
EVALUATE Topic			
→ Summarize			12
→ Summarize			12a
→ Summarize			12b
→ Summarize			12c

Note: The table is a short-hand notation of the rules. The main functions are only mentioned once.

The TRACE table given above describes which information the writer has Selected about André Agassi, in what order, and what he has Reflected on.

The writer *introduced* the topic by mentioning his name and a saliency without any further evaluations (segment 1). After this introduction, the writer *chose* several themes to discuss, and presented aspects about Personal Particulars (2), Career (3), Appearance (6), Character (7) and Hobby (9), in that order. He *elaborated* on each of these presented aspects (2a, 3a, 4, 5, 8, 10, 11), except for the one about Agassi's Appearance. The writer also added a context to some aspects in the form of background information (5a, 7a) and a condition (10a). Once, in elaborating on Agassi's Character, he *evaluated* a specific characteristic (8). Finally, the writer concluded by summarizing Agassi's main characteristics (12-12c).

As this example illustrates, applying the TRACE rules to a text and its structure yields a description of aspects of Reflect and Select operations as executed by the writer.[22] The next section discusses the rate of success in applying these rules to the corpus analysis in order to reveal special instances of the conceptual activities of the subjects investigated in this study.

5.5 Evaluation of the TRACE model

5.5.1 Goodness of fit

The TRACE model was successfully applied to all PISA structures of the corpus-analysis. The set of rules, presented in tables 3, 4, and 5, appeared adequate in describing the entire corpus in terms of Reflect and Select operations. Two defects, however, were found. The corpus did contain two segments which could not be described by TRACE. One segment indicated an encoding problem (*how do you write that*?), and the other indicated a general wish at the end of the text, which proved to be too general to be regarded as an instantiation of EVALUATE (Topic) (see (31)).

(31) Everybody passes through life as he is and later in life external beauty will prove to be of little importance. Yet we must strive to develop the best possible character by following the example of others.
Iedereen geraakt door het leven zoals hij is en later zal de uiterlijke schoonheid toch niet van belang zijn. Toch moeten wij ervoor zorgen door het voorbeeld van anderen een zo goed mogelijk karakter na te streven.

The model can be regarded as a stereotype set of rules for describing a topic. Distinctions between writers result from the *use* of different rules (for instance, not evaluating the topic at

the beginning as is the case in the example), in a different *frequency* (for instance, the rules CHOOSE THEME and ELABORATE THEME), and in a different *order* (for instance, Appearance-Character-Hobby versus Hobby-Character-Appearance).

Applying the TRACE rules to the corpus reveals two types of deviations from the regularities mentioned above: the order in which the functions are applied, and the uniqueness of the topic described. These small deviations, indicated by TRACE, can be described in extenso. With respect to the order of functions, it appeared that

- two writers used an event to introduce a person (see (32) and (33));
- four writers did not mention the name of the topic at the beginning of the description, but only at the end. In one case, it was clearly a rhetorical strategy used to pique interest (see (34));
- in one case, the writer did not discuss the assignment at the beginning but at the end in order to reflect on the non-fulfilment of the assignment (see (35)).

(32) *During the war large numbers of bombing raids were carried out by the Germans. But Dutch boys also did their utmost to destroy as much German property as possible. This was often achieved using airplanes.* Hans Swarts was one of those boys.

In de oorlog werd er veel gebombardeerd door de Duitsers. Maar ook Nederlandse jongens deden hun best om zoveel mogelijk van de Duitsers te vernietigen. Dat gebeurde dikwijls uit vliegtuigen. Hans Swarts deed daar ook aan mee.

(33) *The world of football plays an important role for many players all over the world. For most of them, it is their profession and they have to earn a living with, for example, their speed and tactics* as Lew Yashin does.

De voetbalwereld van vandaag speelt bij vele spelers op de wereld een belangrijke rol. Voor de meeste van hen is dat hun beroep en moeten door bijv. snelheid tot tactiek hun brood verdienen, zoals Lew Yashin dat ook doet.

(34) *Well*, I would like to be in Julia Roberts's shoes.

Tja, ik zou best weleens in de huid van Julia Roberts willen kruipen.

(35) *I have not completely answered the question, I'm sorry, but it was rather difficult.*

Ik heb niet helemaal geantwoord op de vraag, excuseert u mij, maar het ging moeilijk.

With regard to the uniqueness of the topic it appeared that eight writers not only described their ideal person, but aspects of themselves as well. Four of them interspersed their

description with their own characteristics (see (36) and (37)). The total number of such segments occurring in the corpus was 16.
The four other writers made a complete topic-switch from describing a (constructed) ideal to an ideal ego, indicated by an irrealis in which the person described may be regarded as internalizing the ideal topic (see (38)).

(36) Dorus can tell good jokes, but he often appears on tv at 10 pm. and *that is why we are hardly ever allowed to watch. Mother thinks it is too late.*
Dorus kan goed grappen maken, maar hij komt dikwijls om 10 uur op de televisie, en daarom mogen wij haast nooit kijken. Dat vindt moeder te laat.

(37) He lives in Millingen aan de Rijn. *I have lived there too.*
Hij woont in Millingen aan de Rijn. Ik heb daar ook gewoond.

(38) I would like to be like a singer, who travels around the world with his mates, who play for t.v. and radio, and who earn a lot of money but do not misuse it, but, for instance, give their money to other people, who have much more difficulty in earning it *than I have. I would really feel like singing in order not to make myself happy but them (the poor).*
Ik zou willen gelijken op een zanger, die met zijn mannen alle dagen de wereld doorkruisen, die voor t.v. en radio spelen, en die veel geld verdienen maar er geen misbruik van maken, die bijv. hun geld aan de anderen geven die het veel moeilijker verdienen dan ik. Ik zou werkelijk gaan zingen om niet mij maar hun (dus de armen) gelukkig te maken.

In conclusion, the TRACE model provides a tool for describing aspects of the Reflect and Select operations for each individual writer, by specifying the procedures executed.

5.5.2 Scope and limitations of TRACE

The model provides insight into the procedures writers use in generating descriptive texts. The results of the individual writers can be used for comparing and categorizing (groups of) procedures, which is obviously useful in developmental research. The use of TRACE in exploring and describing developmental aspects is discussed in the next chapter.

The TRACE model for descriptive texts specifies the content of conceptual operations in portraying a person. The generalizibility does not imply that these rules can be applied unambiguously to an arbitrary corpus of descriptive texts, because the definition of some embedded functions, such as CHOOSE THEME, are too specific for the topic. The value is the structure of the model: a restricted set of rules ordered around two embedded main functions,

Reflect and Select, of which Select is the basic one, and each of which consists of subfunctions including other functions, etc.. This design can be generally applied to different text types. Van Wijk (1995) has shown that the system also holds for argumentative texts.

The TRACE model presented in this chapter indisputably has its limitations. In two respects it is a restricted characterization of the operations.

First, the TRACE model is restricted to the *content* of the operations, and does not describe their temporal course. This limitation is obviously related to the fact that the model presented here is based on off-line data only. By supplying the TRACE model with on-line data, such as pauses, the model probably provides an adequate basis for developing a real-time model of conceptual processes (see also Schilperoord & Van der Pool 1995).

Second, it is not complete in the sense that it does not cover *all* the operations; it includes the instances of Reflect and Select that leave observable traces in the text. The REFLECT operations, for instance, do not distinguish between guide and control as both are on-line activities which cannot be distinguished reliably from a text.

Despite these limitations, TRACE makes it possible to describe inner operations of individual writers. This provides a basis for the developmental research question of this study which will be discussed in the next chapter.

Notes

1. PISA structures can also be used to evaluate the quality of texts, to provide advice on revision (Van Wijk & Sanders 1987; Sanders & Van Wijk 1991; Van der Pool 1993a; see also chapter 7) or to determine writing skills (Van der Pool 1995a).

2. A TRACE model has been developed for argumentative texts (Van Wijk 1995) and descriptive texts (this study).

3. This basic operation can also be extended by Shape operations. As stated previously, this study focuses on Reflect and Select processes.

4. Bereiter and Scardamalia state that the essence of reflection in writing is constituted by an interaction between the two problem spaces, represented by a 'dual-problem space model' (1987: 303).

5. The terms 'global and local organization' and coherence are used to refer to mental representations of the message. Terms such as super-, macro-, and microstructure are used to refer to aspects of the text structure.

6. Strictly speaking, Reflect, Select and Shape operations result in the preverbal message in accordance with the blueprint presented in chapter 2, section 2.3. As Reflect and Select contribute to that preverbal message and only indirectly to the text, the final term of this production is called *Message*, thereby indicating a substantial part of the preverbal message.

7. As stated in chapter 2, Levelt's macroplanning corresponds to what we have called Select. The step of determining communicative intentions, the pre-macroplanning, refers to what we have termed Reflect.

8. This study does not aim at either natural language generation or user modelling.

9. Additional claims of other text types specific to PISA structures , do not detract from the ones mentioned in this table.

10. The traces found for Select and Reflect are not specified for the suboperations, *retrieve- organize*, and *guide-control*, respectively.

11. This cannot always be claimed, because revising a text may lead to a different organization. However, the Select-operations concerning organize are assumed to be the consequences of the order in which the information is retrieved, as the corpus texts are first draft versions (see also chapter 2, section 2.3).

12. In knowledge-telling these decisions follow naturally from the procedures; in knowledge transforming, these decisions are of a problem-solving nature. The writer is aware of different possibilities for continuation, and has to decide which one to choose (= problem). Solving this problem in a reasonable fashion leads to a decision retraceable in the structure. It is supposed that the nature of the decision -in PISA structural terms, the nature of the connection- is the prime indicator of the type of operation, that is, whether it will be a Select or a Reflect operation.

13. Obviously revealing the kind of information which is related to a theme is a rather circular process, as the content per theme was already defined in the shallow knowledge base in the PISA analysis.

14. The fragments originate in the corpus; the numbers refer to the segments in the texts.

15. These themes correspond to those mentioned in the assignment (see section 3.4, chapter 3).

16. Dependent on the text type, the SELECT and REFLECT functions use different subfunctions. TRACE-rules for argumentative texts, for instance, consist of a SELECT function which is rewritten as (ARGUE Topic) and a REFLECT function, which is, in turn, rewritten as EVALUATE (ARGUE Topic) (See Van Wijk 1995).

17. Of course, this implies that the writer has already decided on the topic. In the case of this study, the task restricted the possibilities: the writer had to describe either a concrete or a constructed person he would have liked to be like. This decision was the starting point for activating knowledge and stating it.

18. In our opinion, a person can also be stereotyped by his profession. Thus, an 'actor' may stereotype a famous person, who is probably not likely to be introverted but will possess and use expressive talents. A policeman may be a stereotype of a brave man trying to control evil and maintain justice.

19. Clearly, the temporal organization of the conceptual operations is not reconstructed. No claims are made as to whether ADD-CONTEXT precedes or follows CHOOSE THEME with respect to the on-line operations. As TRACE is a *text-based* reconstruction, only the PISA structures display the result of the conceptual

operations as being the ADD-CONTEXT prior to the CHOOSE THEME operation. However, as was emphasized before, no claims are made about the real-time aspect of the operations.

20. See, for instance, Hovy 1993 (presentational relations).

21. These events do not necessarily elaborate on segments, they may also elaborate on a theme. In terms of the PISA structure, this means that events may be subordinated since they elaborate on a segment which describes a characteristic, or they may be coordinated to the Characteristics line since they illustrate a theme which has been discussed.

22. The moment at which the operations take place or have taken place is irrelevant in this study. As this reconstruction is based solely on off-line data (that is, PISA structures) no claims are made about the real time course of the operations. This is one of the issues which may prove interesting for further research (see Chapter 7).

6 TRACE and developmental aspects of conceptual processes

6.1 Introduction

The previous chapter described TRACE as a method for modelling aspects of Reflect and Select in so far as they leave observable traces in a text. These instances of conceptual processing can be reconstructed for every writer, on the basis of the texts produced. All of these individual reconstructions, in turn, can be used to gain insight into developmental aspects.

In this chapter, it is examined how developmental aspects can be described by comparing subjects from different age groups. Thus, the focus shifts from the conceptual processes, in research question 2a, to the *development* of these conceptual processes, as formulated in research question 2b. The characterization of developmental changes in terms of TRACE is the main issue of this chapter (see the marked box in figure 1).

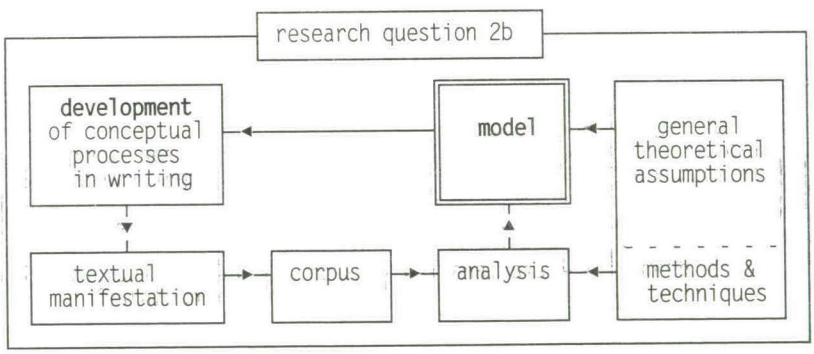


Figure 1. The position of this chapter in the study

TRACE rules reveal how writers presumably operated, in other words, the TRACE rules can be regarded as the *procedures* executed by the writer. These procedures have been used to characterize individual performances. These characterizations reveal different patterns of application of the procedures by different groups of writers. These regular patterns of procedures are directed by *strategies* (see chapter 2, section 2.3). These strategies can be used in modelling the development of the conceptual processes.

On the basis of the TRACE rules for describing a person, two main strategies were postulated. These strategies are described in section 6.2. Their relation to age and choice of topic is discussed in section 6.3. The chapter is concluded with a discussion of the scope and limitations of TRACE in describing development (section 6.4).

6.2 Writing strategies in terms of TRACE

Applying the TRACE model to the corpus texts resulted in 86 unique TRACE tables, one for every text. The differences between the reconstructions were primarily due to differences in the order in which the procedures were applied (see also chapter 5). By disregarding the precise order in which the TRACE *procedures* needed to be applied in order to reconstruct the conceptual activities, it was possible to extract equivalent sets of applied rules. This clustering resulted in two distinctive methods of writing, which were regarded as the leading *strategies* in writing a description of a person. They are represented in (1).

(1) **Strategies of the writer**
   a. Describe strategy: Present information about topic X
   b. Evaluate strategy: Reflect on task and information presented about topic X

The Describe strategy yields texts that can be reconstructed by the procedures of (DESCRIBE Topic) only. The Evaluate strategy yields texts that can be reconstructed by (EVALUATE (DESCRIBE Topic)). These two strategies are cumulative in the sense that the Evaluate strategy always implies that the Describe strategy will be executed (see table 1). Note that the procedures are written in capitals and the strategies in lower case.

Table 1. Relation between strategies and procedures

Strategy	Procedures within strategy	
Describe		DESCRIBE Topic
Evaluate	EVALUATE	(DESCRIBE Topic)

Within the two strategies of increasing complexity variants can be distinguished. In the present discussion, only the extremes of this continuum will be dealt with: the simple and the

complex. Both strategies will be discussed in the next two subsections; in section 6.2.1 the Describe strategy is addressed in its simple and complex form, in section 6.2.2 the Evaluate strategy in its simple and complex form.

6.2.1 The Describe strategy

The Describe strategy leads to a factual description without any subjective comments. The *simple* form of this strategy and the more *complex* method of composing a description will be discussed. These two subtypes are characterized by the procedures presented in table 2.

Table 2. Variant within the Describe-strategy represented by the procedures

Describe-strategy	Procedures		
simple variant	INTRODUCE Topic	→	Name or Stereotype
	CHARACTERIZE Topic	→	CHOOSE THEME
complex variant	INTRODUCE Topic	→	Name or Stereotype + **Saliency**
	CHARACTERIZE Topic	→	**ADD CONTEXT**
			+ CHOOSE THEME
			+ **ELABORATE THEME**

The *simple* method of description consists of mentioning the topic by its name, or stereotyping it. Subsequently, the topic is characterized through the discussion of several themes. The *complex* form differs from the simple form in that the discussion of the topic and the characterization are more extensive. A saliency is added to the name or stereotype in introducing the topic and contexts and/or elaborations are added to the characterization. Both types of writing are illustrated as TRACE reconstructions of a text.

The simple variant

An example of a TRACE reconstruction in which the most simple variant of the Describe strategy is displayed is presented in table 3; the writer is a 10-year-old boy. On the left-hand side, the TRACE rules are presented, on the right-hand side, the results in the form of text segments (including spelling errors).[1]

Table 3. A TRACE Reconstruction of conceptual activities corresponding to the *simple variant* of the *Describe strategy*. The segments of the text form the output[2]

TRACE rule		Segment	
DESCRIBE Topic → **INTRODUCTION Topic** + **CHARACTERIZE Topic**			
INTRODUCTION Topic → name		1	gerad van Tienen.
CHARACTERIZE Topic*			
→ CHOOSE THEME →	Personal Particulars	2	8 years old
→ CHOOSE THEME →	Appearance	3	he looks well.
→ CHOOSE THEME →	Career	4	his learning is prety average.
→ CHOOSE THEME →	Personal Particulars	5	his father is a Carpenter.
→ CHOOSE THEME →	Hobbies	6	he spends much time outside.
→ CHOOSE THEME →	Hobbies	7	he plays a lot.
→ CHOOSE THEME →	Personal Particulars	8	eats little.
→ CHOOSE THEME →	Personal particulars	9	he makes himself dirty.
→ CHOOSE THEME →	Hobbies	10	often with other people.
→ CHOOSE THEME →	Appearance	11	small.

Key: '*' recursive use of procedure

This reconstruction makes explicit the manner in which the writer has described the person he would have liked to resemble by repeatedly choosing a theme and presenting an aspect. After the presentation of an aspect, another theme is chosen, another related aspect is presented, etc.. The repetition of the themes of Appearance, Hobbies, and Personal particulars indicates that this writer did not retrieve information very systematically. He first discusses a Personal particular of the topic, followed by an aspect of the topic's Appearance, his Career, and again something about his Personal particulars. Then he discusses two aspects of the topic's Hobbies and Personal particulars. And finally, the writer again retrieves information about the topic's Hobbies and his Appearance.

This example represents the most elementary, simple manner of describing a person.[3] The writer has taken care of *topic coherence* by presenting aspects directly related to the person he would have liked to resemble. All information selected is immediately linked to the topic.

The complex variant

The complex manner of describing includes extensions with respect to the introduction of the topic (*Saliency*,) as well as with respect to its characterization (ADD CONTEXT, ELABORATE). This extended set of rules makes possible the composition of a hierarchically more complex

text, in particular with regard to the characterization of the topic. The presented aspects may be contextually embedded by an added context remark (see chapter 5 for examples). This may consist of a background or condition to the characteristic or a background to a theme in the form of an explicit theme marker. The writer may also elaborate a theme by specifying a characteristic or presenting an event. An example of a TRACE reconstruction of this complex variant of the Describe strategy is presented in table 4; the writer is a 12-year-old boy.

Table 4. A TRACE Reconstruction of conceptual activities corresponding to the *complex variant* within the *Describe strategy*. The segments of the text form the output[4]

TRACE rule			Segment	
DESCRIBE Topic				
→ **INTRODUCTION Topic + CHARACTERIZE Topic**				
INTRODUCTION Topic				
→ name + salient feature			1	E.P. 's Graafland. Football player
CHARACTERIZE Topic*				
→ CHOOSE THEME	→	Career	2	He is a professional football player with Fijenoord.
→ ELABORATE	→	Specify	3	Keeper
→ ELABORATE	→	Specify	4	He is a good goal-keeper
→ CHOOSE THEME	→	Appearance	5	He has black hair
→ CHOOSE THEME	→	Appearance	6	and is about 1.69 m.
→ ADD CONTEXT	→	Condition	7a	If the players are having a quarrel
→ CHOOSE THEME	→	Character	7	he intervenes
→ ELABORATE	→	Consequence	8	and it is finished in a minute
→ ELABORATE	→	Event	9	During the match against real Madrid in Holland there was a lot of rough playing
→ ELABORATE	→	Event	10	A Spaniard was tripped up
→ ELABORATE	→	Event	11	and then it started
→ ELABORATE	→	Event	12	On the stands someone had hit somebody over the head with a beer bottle.
→ CHOOSE THEME	→	Character	13	He is kind
→ CHOOSE THEME	→	Career	14	and a good goal-keeper
→ ELABORATE	→	Specify	15	He has number 1.

This writer, in contrast to the previous one, has not only established *topic coherence* by presenting theme-related aspects, resulting in the segments 2, 5, 6, 7, 13, and 14, but he has also created *local coherence* by adding contexts and elaborations to the presented characteristics.

Almost every time the writer has chosen a theme, he elaborates it by specifying the characteristic presented. The writer, for instance, elaborates on the career of the protagonist (segment 2) by specifying that Graafland is a keeper (segment 3), even a good one (segment 4). He also illustrates - although in an implicit and incomplete way - a character trait by telling about an event. Graafland's 'intervening strength' is exemplified by means of an event which had taken place, 'the incident with the beer bottle'.

6.2.2 The Evaluate strategy

Writers using the Evaluate strategy compose a description interspersed with evaluative remarks. Parallel to the Describe strategy, within this type, two variants of increasing complexity can be distinguished as well: a simple one and a complex one. The subtypes are characterized by the procedures executed in table 5.

Table 5. Variants within the Evaluate strategy represented by the procedures

Evaluate-Strategy	Procedures	
simple variant	INTRODUCE Topic	→ Name or Stereotype + [Saliency]
	EVALUATE (CHARACTERIZE Topic)	→ CHARACTERIZE Topic + [GENERAL-EVALUATION (CHARACTERIZE Topic)] + [SPECIFIC-EVALUATION (CHARACTERIZE Topic)]
complex variant	EVALUATE (INTRODUCE Topic)	→ INTRODUCE Topic + **[EVALUATE (INTRODUCE)]** + **[EVALUATE (Topic)]**
	EVALUATE (CHARACTERIZE Topic)	→ CHARACTERIZE Topic + [GENERAL-EVALUATION (CHARACTERIZE Topic)] + [SPECIFIC-EVALUATION (CHARACTERIZE Topic)]
	EVALUATE (Topic and/or Description)	→ **[EVALUATE (Topic)]** + **[EVALUATE (Description)]**

The *simple* way to evaluate a description is by concentrating the reflection on the characterization of the topic, that is by making specific or general evaluative remarks with

respect to the traits presented (GENERAL-EVALUATION (CHARACTERIZE Topic), SPECIFIC-EVALUATION (CHARACTERIZE Topic)]). The *complex* way adds to these evaluations both a reflection when introducing a topic (EVALUATE (INTRODUCE), EVALUATE (Topic)) and a reflection when concluding the description (EVALUATE (Topic), EVALUATE (Description)), printed in bold in table 5.

The simple variant

The simple variant within the Evaluate strategy includes the same rules as the Describe strategy in its complex form, with this difference that the Evaluate strategy is supplemented with procedures for an evaluation on the characterization of a general or specific kind. An example of a text which complies with this variant of the Evaluate strategy was presented in chapter 5 (see table 7).

The complex variant

The complex variant consists of an extension of evaluations at the beginning of the description (EVALUATE (INTRODUCTION Topic)) and at the end (EVALUATE (Topic and/or Description)). An example of a reconstruction based on this set of procedures is presented in table 6; the writer is a 15-year-old boy.

The reconstruction in table 6 shows that the writer has not only evaluated the characterization of the topic (EVALUATE (CHARACTERIZE Topic)), but also the introduction of the topic (EVALUATE (INTRODUCTION)) and the description (EVALUATE (Description)). The writer reflects on the mentioning of the topic's name by showing that he wants to keep him anonymous. In order to avoid a problem of reference in the continuation of the text, he makes up a name (segment 2). The continuation of the text, the characterization, is interspersed with evaluations, opinions (segments 5 and 6) or meta-textual remarks (segment 10). The opinions which the writer expresses are not by default positive, see, for instance, segment 21b, which implies a critical attitude towards the person described. The meta-textual remarks, such as segments 14 to 16, show an awareness of a reader (see segment 16) and of the preceding text (see segment 20a). The writer concludes his composition with a reflection on the entire description (EVALUATE (Description)). With such a conclusion, the writer consciously finishes his text and realizes - in relation with the explicit introduction - a global frame for the text.

Table 6. A TRACE Reconstruction of conceptual activities corresponding to the *complex variant* within the *Evaluate strategy*. The segments of the text form the output[5]

TRACE rule		Segment	
EVALUATE (DESCRIBE Topic)			
→ **EVALUATE (INTRODUCE Topic)**			
+ **EVALUATE (CHARACTERIZE Topic)**			
+ **EVALUATE (Topic and/or Description)**			
EVALUATE (INTRODUCE Topic)*			
→ INTRODUCE Topic	→ Saliency	0	A gym teacher
	→ Saliency	1	The person I am going to describe teaches at our Ulo school
→ EVALUATE (INTRODUCE)		2	and to make things easier
→ INTRODUCE Topic	→ Name		we will call him John
EVALUATE (CHARACTERIZE Topic)*			
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→ Character	3	John is a very easy person to get on with,
→ ELABORATE	→ Specify	3a	who does not punish you,
→ ELABORATE	→ Condition	3b	that is if you are not 'asking' for it.
→ CHOOSE THEME	→ Career	4	He is a very good teacher
→ ELABORATE	→ Specify	4a	who teaches gymnastics in a nice way,
→ ELABORATE	→ Exception	4b	although it is sometimes boring.
→ GENERAL-EVALUATION	→ Positive	5	By the way I can get on very well with him
→ SPECIFIC-EVALUATION	→ Positive	6	And I also intended to learn the profession that John teaches.
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→ Personal part.	7	I do not know John's age
→ ELABORATE	→ Specify	8	but I take him to be about 30
→ SPECIFIC-EVALUATION	→ Positive	8a	which, by the way, I think is a nice age for teaching those lessons.
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→ Character	9	John's character is unknown to me,
→ SPECIFIC-EVALUATION	→ Meta-textual	10	but I will try to write something about it
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→ Character	11	In my view he has a calm character
→ ELABORATE	→ Reason	12	for he does not get excited very easily
→ CHOOSE THEME	→ Personal part.	13	I also think that he has a good marriage with one child.
→ SPECIFIC-EVALUATION	→ Meta-textual	14	That is all I can say,
→ SPECIFIC-EVALUATION	→ Meta-textual	15	I can guess of course
→ SPECIFIC-EVALUATION	→ Meta-textual	16	but that is of not much use to the person who examines this text I think.
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→ Appearance	17	John's figure is good, in my view
→ ELABORATE	→ Specify	18	the length of legs, arms, trunk and lower part of the body is good,

→ ELABORATE	→ Specify	19 and the face with the blue eyes and the black hair is not bad either.
→ CHOOSE THEME	→ Career	20 John is a gym teacher
–/ SPECIFIC-EVALUATION	→ Meta-textual	20a as was already mentioned in some lines at the beginning
→ CHARACTERIZE Topic		
→ CHOOSE THEME	→ Hobbies	21 In my view he seeks relaxation in the garden and in the living room (or outside)
→ ELABORATE	→ Specify	21a where he reads some book or other
→ ELABORATE	→ Specify	21 and in riding his scooter
→ SPECIFIC-EVALUATION	→ Negative	21b (which by the way I think complete nonsense)
EVALUATE (Description)	→ Metastatement	22 This is my story about the gym teacher of our school.

On the basis of the above characterizations of writing strategies, the second research question of this study can be specified in two subquestions:

a. Do the strategies applied by our subjects correlate with age (10, 12, 15 year olds and adults)?
b. Do the strategies correlate with the choice of the topic (concrete versus constructed topic)?

The first question concerns the developmental aspects related to age. This is discussed in section 6.3 by comparing the four age groups all describing a concrete topic (10-, 12-, 15-year-olds and adults); in section 6.3.1 the results are presented for the development within the age groups of 10-, 12- and 15-year olds, section 6.3.2 for the development from 15 year olds to adulthood. The second question, the relation with the choice of the topic, is central in section 6.4. To answer this question, the two groups of 15-year-old subjects both describing a different kind of topic are compared (the ULO pupils described a concrete topic, the Grammar school pupils a constructed one).

6.3 Writing strategies in relation to age

In order to examine the influence of age controlled for the choice of the topic only, the writers describing a concrete person were considered.[6] They were of four age groups: 10-, 12-, and 15-year-olds and adults. Table 7 presents the distribution of these four age groups over the two main strategies, the Describe strategy and the Evaluate strategy.

Table 7. Distribution of the Describe and Evaluate strategy over the age groups (in proportions)

	Describe	Evaluate	Describe		Evaluate	
			simple	complex	simple	complex
10 years (n=10)	.90	.10	.10	.80	.00	.10
12 years (n=20)	.65	.35	.00	.65	.05	.30
15 years (n=20)	.30	.70	.00	.30	.15	.55
adults (n=12)	.00	1.00	.00	.00	.08	.92

All results were tested with non-parametric tests (the Kruskall-Wallis H test for 3 or more independent sample; the Mann-Whitney ranksum test for 2 independent samples).
The distribution displays a relation between age and the occurrence of the two strategies. As writers get older, the Describe strategy is applied less often, a decrease from .90 to .00 in favour of the Evaluate strategy ($\chi^2(3)= 23.04$, $p<.001$).

A division within the strategies into the variants *simple* and *complex* learns that the simple form of the Describe strategy is applied by a single 10-year-old writer. The simple variant within the Evaluate strategy is also applied far less often than the complex one (varying from .00 to .15). In both strategies, the majority of the writers used the complex variant.

6.3.1 Development within the age groups 10, 12, and 15-years old

Strategies and procedures
The shift from the Describe strategy to the Evaluate strategy means an extension in the procedures applied. Whereas in the Describe strategy only the set of DESCRIBE Topic procedures is used, in the Evaluate strategy, this set is extended by EVALUATE procedures.

There are no strong expectations about the development, but one could raise the question whether the execution of DESCRIBE procedures differs for the two strategies. One could argue that the Evaluate strategy influences the execution of the DESCRIBE procedures in two ways, both related to the frequency of applying the procedures within DESCRIBE Topic. First, if the Evaluate strategy leads to a more critical attitude towards the relevance of the information retrieved, evaluation may result into a decrease in the use of CHOOSE THEME procedure. Second, it may result in an increase of applying ADD CONTEXT and/or ELABORATE. If the evaluative way of writing does not stay close to bare facts but aims at presenting the information in a nuanced way.

In order to check for the influence of the Evaluate strategy on the execution of the DESCRIBE procedures, the frequency of using the procedures CHOOSE THEME and ADD CONTEXT

plus ELABORATE were examined. As the procedures ADD CONTEXT and ELABORATE do not occur within the Describe strategy in its simple form, the comparison is made for the Describe strategy in its complex from on the one hand, and the Evaluation strategy (including both the simple and complex variant) on the other hand. In addition, the adults are excluded from the comparison as they always applied the DESCRIBE procedures from the governing Evaluate strategy.

The mean frequency of using the procedure CHOOSE THEME, on the one hand, and ADD CONTEXT and ELABORATE, on the other, is presented in table 8 for the three age groups 10, 12, and 15 years old. (The age groups have been divided into subjects applying the Describe strategy, indicated by n1, and those using the Evaluate strategy, indicated by n2.)

Table 8. Application of the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE in relation to strategy and age (in average numbers over subjects)

		CHOOSE THEME		ADD CONTEXT/ELABORATE	
		Describe	Evaluate	Describe	Evaluate
10 years	(n1=9, n2=1)	6.00	6.00	7.11	7.00
12 years	(n1=13, n2=7)	8.92	10.14	10.00	14.29
15 years	(n1=6, n2=14)	11.67	12.29	20.00	21.79

There was no relation between strategy and the frequency of applying the procedure CHOOSE THEME (for every age group $z<1.5$, n.s.). The same held for the procedures ADD CONTEXT/ELABORATE (for every age group $z<1.5$, n.s.).

Thus, the Evaluate strategy does not affect the *frequency* of applying the procedures within CHARACTERIZE Topic. One could expect, however, that the Evaluate strategy influenced the *order* in which the themes (Appearance, Career, etc.) were chosen in applying the procedure CHOOSE THEME. One could argue that the Evaluate strategy influenced the thematical continuity in the sense that once a theme has been chosen, it is described exhaustively. Such a way of writing reduces the necessity of choosing a theme for the second time in order to add information. This would show up in a smaller number of discontinuities in applying the procedure CHOOSE THEME within the Evaluate strategy.[7]

To test this conjecture, the number of times a theme is re-chosen a second time is informative. The number of thematic discontinuities in applying the procedure CHOOSE THEME are presented for both strategies in table 9.

Table 9. Discontinuous application of the procedure CHOOSE THEME in relation to strategy and age (in average numbers over subjects)

	Describe	Evaluate
10 years (n1=9, n2=1)	1.78	1.00
12 years (n1=13, n2=7)	1.77	2.14
15 years (n1=6, n2=14)	2.33	2.57

The results were tested with the Mann-Whitney rank sum test. Within the age groups, there was no difference between the two strategies (for every age group, $z<.65$, n.s.). This means that within the Describe-strategy as well as within the Evaluate strategy, writers of the same age produce a comparable number of discontinuities.[8] The Evaluate strategy did not affect the thematical coherence positively.

A related question is: How has the information been distributed over the first and second mentionings of themes.[9] Is it the case that under influence of the Evaluate strategy the number of segments in the second mentioning of a chosen theme is smaller than for the Describe strategy, because most of the information has already been presented the first time?

In order to answer this question, the number of segments in these second mentionings establishing the discontinuity is contrasted with the number of segments in the first mentioning of a theme and related to the strategies (see table 10).

Table 10. Number of segments in the first and second mentionings of themes in relation to strategy and age (in average numbers over subjects)

	First mentioning		Second mentioning	
	Describe	Evaluate	Describe	Evaluate
10 years (n1=9, n2=1)	9.44	9.00	3.67	4.00
12 years (n1=13, n2=7)	12.54	18.14	6.38	6.29
15 years (n1=6, n2=14)	22.00	19.29	9.67	14.79

None of the first mentionings differed in the number of segments for the two strategies (for every age group $z< 1.60$, n.s.). Second mentionings showed no relation with strategy either (for every age group $z< .60$, n.s.). Thus, in both strategies, writers of the same age present more or less the same amount of information in the first and second mentionings of themes.

Age

The above results show that *within* age groups the DESCRIBE procedures are applied in the same way with respect to their frequency and thematic order irrespective of the presence of the Evaluate strategy. But how do these features change with age? Does the application of procedures differ for young and older writers? Do 10-year-old writers, for instance, use procedures less frequently than 15-year-old ones? Or do they have more discontinuities in their thematical presentation?

In testing these and other differences *between* the age groups, the strategies are lumped together given the above results.

First, we will examine the frequency of applying the DESCRIBE procedures CHOOSE THEME and ADD CONTEXT/ELABORATE for the three age groups irrespective of the strategy used; subsequently, the discontinuous application of the procedure CHOOSE THEME.

Table 11 shows the frequency of applying the procedure CHOOSE THEME and ADD CONTEXT and/or ELABORATE for the age groups 10-, 12- and 15-years.

Table 11. Application of the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE in relation to the age (in absolute numbers)

	10 years	12 years	15 years
CHOOSE THEME	6.00	9.35	12.10
ADD CONTEXT/ELABORATE	7.10	11.50	21.25

The frequency of applying the procedure CHOOSE THEME showed a significant relation with age ($\chi^2(2)=19.12$, $p<.001$). The older the writer the more this procedure was applied; 15-year-old writers apply this procedure twice as often as 10-year-olds.

The application of the procedures ADD CONTEXT and ELABORATE also showed a significant increase with age ($\chi^2(2)=16.76$, $p<.001$). Older writers add three times more contexts and elaborations to the characteristics presented than the 10-year-olds.

The increase in the use of the procedures is reflected in the number of different themes discussed as well as the number of segments per theme.

First, as writers get older, they describe more different themes. Whereas 10-year-old writers describe on the average 3.5 themes, 12-year-olds describe 4.4 themes, and 15-year-old writers 4.7 ($\chi^2(2)=13.02$, $p<.005$).

Second, an older writer presents per theme more information, in number of segments. Four themes showed a regular increase with age (Character $\chi^2(2)=15.69$, $p<.01$; Appearance $\chi^2(2)=5.90$, $p<.05$; Career $\chi^2(2)=13.90$, $p<.01$; Hobby $\chi^2(2)=.94$, $p=.66$). Only Personal particulars showed an irregular pattern ($\chi^2(2)=6.41$, $p<.05$). The 10-year-olds presented more information than the 12-year-old writers, but less than the 15-year-olds. In contrast to the older subjects, the 10-year-olds also described idiosyncrasies, such as nail biting.

Thus, older writers present more information, but how do they order that information? Do they present aspects related to one and the same theme at once, in other words, do they have fewer discontinuities? Or do they present the same amount of information as younger ones in discussing a theme the first time, and does their development appear from adding more information the second time?

The number of times the procedure CHOOSE THEME is discontinuously applied and the numbers of segments in the first and second mentionings of themes are presented in table 12.

Table 12. Discontinuous application of the procedure CHOOSE THEME and the number of segments in the discontinuities in relation to age (in absolute numbers and proportions)

	10 years	12 years	15 years
Number of discontinuities	1.70	1.90	2.50
Proportion of themes reoccurring	.46	.42	.53
Number of segments in first mentioning	9.40	14.50	20.10
Number of segments in second mentioning	3.70	6.35	13.25
Proportion of segments in second mentioning	.31	.30	.37

Age groups did not differ much in the number of discontinuities ($\chi^2(2)=2.94$, $p=.23$) nor in the proportion of themes reoccurring ($\chi^2(2)=1.43$, $p=.49$; the number of themes reoccurring is expressed in the number of discontinuities with respect to the number of different themes that are discussed). Thus, older writers do not differ much from younger ones in number and proportions of reintroducing themes.

For the distribution of the information over the first and second mentioning different results are found. The number of segments in the first and second mentionings both increased

significantly with age (first mentioning $\chi^2(2)=11.49$, $p<.01$; second mentioning $\chi^2(2)=8.23$, $p<.05$). The older a writer is, the more he discusses the first time as well as the second time he mentions a theme. The proportion of segments in the second mentioning did not change very much ($\chi^2(2)=.52$, $p=.77$).

The above results of the development with age showed an increase for the number of times the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE were applied. This increase is reflected in the first as well as in the second mentionings of themes; the proportion is, however, almost equal for all age groups. The 10-, 12- and 15- year-old writers did not differ much for the discontinuity of the order in which the themes are discussed.

6.3.2 Development from 15-years of age to adulthood

The results presented in table 7 showed that the main developmental shift takes place between the age of 12 and 15 years. In the previous section the development for the age levels 10, 12 and 15 years of age was examined. This section is devoted to explore the development from 15 years of age to adulthood.

The question to be addressed now is, how the application of the procedures develops from 15-year-old to adulthood. To explore this, we will first examine the frequency of applying the DESCRIBE procedures CHOOSE THEME and ADD CONTEXT/ELABORATE; subsequently, the number and the number of segments in the discontinuous application of the procedure CHOOSE THEME. In addition, the specific EVALUATE procedures are examined.

The frequencies in which the adults applied the procedures CHOOSE THEME and ADD CONTEXT and/or ELABORATE are contrasted with the frequencies for the 15-year-olds (see table 13).

Table 13. Application of the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE in relation to age (in absolute numbers)

	15 years (n= 14)	adult (n= 12)
CHOOSE THEME	12.29	6.17
ADD CONTEXT + ELABORATE	21.79	11.75

In comparison with 15-year-old writers, the adults less often choose themes, a halving of the number of applications (CHOOSE THEME $z=3.40$, $p<.001$). The adults also extended their

descriptions far less than the 15-year-olds (ADD CONTEXT/ELABORATE $z=2.84$, $p<.005$). The adult's number of added contexts and elaborations is half of those added by the younger writers. This decrease is also reflected in the decreasing number of themes (adults 3.67 themes; 15-year-olds 4.71 themes ($z=2.49$, $p<.05$); 15-year-olds presented more information about the protagonist's Career ($z=2.64$, $p<.01$; for the other themes, $z<1.70$, n.s.).

The frequency of applying the procedures within DESCRIBE Topic decreases for adult writers. But what about the thematic order in which adults discuss the themes? One could, for instance, expect that the adults reflect more on the order in which information is retrieved If this conjecture is true, the number of discontinuities would be decreased in adults compositions.

The occurrence of discontinuities in discussing themes and the number of segments in the first mentioning and second mentioning are presented in table 14.

Table 14. Discontinuous application of CHOOSE THEME and the number of discontinuities in relation to age (in absolute numbers and proportions)

	15 years	adult
Number of discontinuities	2.57	.67
Proportion of themes reoccurring	.52	.16
Number of segments in first mentioning	19.29	16.17
Number of segments in second mentioning	14.79	1.75
Proportion of segments in second mentioning	.37	.09

Adults presented a thematically more orderly description of the person they would like to be like. The number of discontinuities and the proportion of themes reoccurring both decrease significantly (number of discontinuities $z=3.14$, $p<.01$; proportion of themes reoccurring $z=10.32$, $p<.01$).

The lengths of discussing themes for the first time did not differ much for the two age groups ($z=.17$, $p=.68$). The number of segments in the second mentionings in the adults texts is significantly smaller than those of the 15-year-old writers, a decrease of about 85% ($z=10.97$, $p<.001$). The proportion of segments in the second mentioning decreased from 37% to 9% with age ($z=3.02$; $p<.01$).

These results indicate that adults behave differently from 15-year-olds when using the same strategy. This may be caused by the fact that adults apply the Evaluate-strategy more intensively. If they do indeed apply the strategy more intensively, this might not only appear from the orderly presentation but also from an increase of evaluative remarks in the introduction (EVALUATE (INTRODUCE/Topic)), in the Description section of the text (GENERAL-SPECIFICATION and/or SPECIFIC-EVALUATION), and remarks that establish the Conclusion EVALUATE (Topic/Description).
The size of these evaluative remarks is presented for both groups in table 15.

Table 15. The number of EVALUATE (INTRODUCE/Topic), GENERAL-EVALUATION and SPECIFIC-EVALUATION in the Description section of the text, and of EVALUATE (Topic/Description) in relation to age

	15 years (n=14)	adult (n=12)
EVALUATE (INTRODUCE Topic)	.64	1.92
GENERAL/SPECIFIC-EVALUATION	2.07	1.67
EVALUATE (Topic and/or Description)	1.78	1.83

The difference between adults and 15-year-old writers manifests itself especially in the evaluative remarks in introducing a topic. Adults tended to evaluate more at the beginning of their composition ($z=1.58$, $p=.11$).

There are no differences in the number of general and/or specific evaluations ($z=.69$, $p=.49$) Nor was there a difference in evaluating the topic or description at the end of the generation process ($z=.26$, $p=.79$).

6.4 Writing strategies in relation to the choice of topic

In order to study the influence of the choice of the topic (a concrete versus a constructed person) on the use of strategies (question b), the pupils describing a constructed person (the grammar school pupils, 15/16 years) are compared to those who characterized a concrete person (ULO-pupils, 15-years). The question is whether the choice of the topic made any difference in using the strategies and applying the procedures.
The distribution of the strategies in relation to topic-choice is presented in table 16.

Table 16. Distribution of the Describe and Evaluate strategies in relation to the choice of topic (in proportions)

		Describe	Evaluate
concrete topic	(n=20)	.30	.70
constructed topic	(n=24)	.54	.46

In producing texts about a constructed topic, the writers tended to use Describe strategy more often (54 versus 30 percent; $z=1.60$, $p=.055$).

The examination of developmental trends for texts about concrete topics showed that until the age of 15 there was no effect of the strategy on the frequency of applying the procedures CHOOSE THEME and ADD CONTEXT and/or ELABORATE (table 8). It also showed that there was no effect on the number of discontinuous applications of CHOOSE THEME (table 8), nor on the number of segments in the first and second mentionings (table 10). The question now is whether writers will differ when describing a constructed topic.

Parallel to the other examinations, we will first look at the frequency of applying the CHOOSE THEME and ADD CONTEXT/ELABORATE procedures in describing a constructed versus a concrete person, and subsequently, at the number of discontinuities and the number of segments in the first and second mentionings, as well as the size of the specific EVALUATE procedures.

The frequencies of applying CHOOSE THEME and ADD CONTEXT and/or ELABORATE are presented in table 17.

Table 17. Application of the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE in relation to the choice of topic and strategy (in absolute numbers)

	CHOOSE THEME		ADD CONTEXT/ELABORATE	
	Describe	Evaluate	Describe	Evaluate
concrete topic	11.67	12.29	20.00	21.79
constructed topic	8.07	7.09	3.23	6.36

The pupils describing a constructed topic applied the procedures less frequently than the pupils describing a concrete topic. For the application of the procedure CHOOSE THEME this difference is significant within the Evaluate strategy ($z=3.13$, $p<.01$) but not for the Describe strategy

($z=1.37$, $p=.18$). The procedures ADD CONTEXT and ELABORATE were used far less for the constructed topic than for the concrete topic in both strategies (Describe strategy $z=3.02$, $p<.01$; Evaluate strategy $z=3.75$, $p<.001$).

In addition, the writers describing a constructed person used significantly fewer different themes. Within the Describe strategy, the number of different themes for a concrete topic is 4.83, whereas the descriptions of a constructed topic cover 3.69 themes ($z=2.16$; $p<.05$). Within the Evaluate strategy the concrete topic is described on 4.79 different themes, the constructed topic on 3.09 different themes ($z= 3.31$, $p<.001$).

The relative size of the themes within the Describe strategy differed significantly for the theme Personal particulars. The texts about a concrete topic comprised more information related to the theme Personal particulars ($z=2.23$, $p<.05$); those about a constructed person, more about the protagonist's Hobby ($z=2.07$, $p=.04$; for the other themes $z<1.60$, n.s.).

The size of themes within the Evaluate strategy differed significantly for the themes Character and Career. The texts about a constructed topic contained more aspects related to the Character of the protagonist ($z=2.42$, $p<.05$); those about a concrete topic included more information about the person's Career ($z=3.68$, $p<.001$; for the other themes $z<1.40$, n.s.).

In addition to the frequencies of applying the procedures, the number of discontinuities in applying CHOOSE THEME has been explored for both strategies. For both strategies, the number of discontinuities in applying the procedure CHOOSE THEME and the number of segments in the first and second mentionings are presented in table 18 in relation to the choice of topic, separately for concrete and constructed topics.

Table 18. Discontinuous application of the procedure CHOOSE THEME and the number of the discontinuities in relation to the choice of topic

	Describe		Evaluate	
	concrete	constructed	concrete	constructed
Number of discontinuities	2.33	.92	2.57	1.18
Proportion of themes reoccurring	.49	.21	.53	.32
Number of segments in first mentioning	22.00	9.69	19.29	10.00
Number of segments in second mentioning	9.67	1.62	14.79	3.45
Proportion of segments in second mentioning	.31	.11	.39	.18

Within both strategies, there were fewer discontinuities in the description of constructed topics (Describe strategy: $z=2.19$, $p<.05$; Evaluate strategy: $z= 2.18$, $p<.05$). The proportion of themes reoccurring was also less for the constructed topics (Describe-strategy: $z=4.98$, $p<.05$); Evaluate-strategy: $z=1.49$, $p=.14$).

The same holds for the number of segments in the first and second mentionings. Within the Describe strategy, the number of segments in the first mentionings for constructed topics was less than half for concrete topics ($z= 2.92$, $p<.01$); the number of segments for the second mentionings decreased by about 75% in the descriptions of a constructed person ($z=2.97$, $p<.01$). Within the Evaluate strategy, the same results are found. The first and second mentionings in the descriptions about a constructed person are also much smaller than those in descriptions of a concrete person (first mentioning: $z=2.30$, $p<.05$; second mentioning: $z= 2.57$, $p<.01$). The proportion of segments in the second mentionings differed as well for both topics within the strategies (Describe strategy: $z=2.46$, $p<.01$; Evaluate strategy: $z=1.87$, $p=.06$).

The above results show significant differences in the choice of topic for the application of the DESCRIBE procedures. What needs to be explored next is whether there are differences for the specific procedures within EVALUATE (DESCRIBE Topic). In table 19, the size of these specific procedures are presented in relation to the choice of topic.

Table 19. The size of EVALUATE (INTRODUCE/Topic), the size of GENERAL-EVALUATION and SPECIFIC-EVALUATION in the Description-part of the text, the size of EVALUATE (Topic/Description) in relation to the choice of topic (in absolute numbers)

	concrete topic (n=14)	constructed topic (n=11)
EVALUATE (INTRODUCE Topic)	.64	.82
GENERAL/SPECIFIC-EVALUATION	2.07	1.73
EVALUATE (Topic and/or Description)	1.79	2.00

It appears that the differences are not significant for any of the procedures (for the evaluative remarks in the introduction: $z=.40$, $p=.69$; for the characterization-part: $z=.62$, $p=.54$.; for the conclusion-part: $z=.14$, $p=.89$). Thus, the evaluative remarks made about a constructed and a concrete person do not differ in a reliable way.

6.5 Conclusion

6.5.1 Summary of the results

The TRACE rules presented in chapter 5 revealed the *procedures* writers presumably applied in describing a person. These procedures were used to characterize the individual performances of the 86 subjects. This yielded different patterns of applying the procedures, which we assumed to be *strategies*. We postulated two main strategies for describing a person: a Describe strategy including the DESCRIBE procedures, and an EVALUATE strategy including EVALUATE (DESCRIBE Topic) procedures. The complexity of these strategies varied on a scale from *simple* to *complex*. Each strategy plus its two extremes (simple and complex) were discussed and illustrated in section 6.2.

We used the two strategies and the associated procedures to reveal developmental aspects of the conceptual processes. The main questions in this chapter were:

a. Do the Describe and Evaluate strategy applied by our subjects correlate with age (10, 12, 15 years of age and adults)?
b. Do the strategies correlate with the choice of the topic (concrete versus constructed topic)?

We will briefly summarize the results in order to answer these questions.

a. Do the Describe- and Evaluate-strategy correlate with age?

The main effect is that there is a relation between age and the occurrence of the strategies. From the age of 15, most writers use the Evaluate-strategy. In both strategies, almost all writers applied the complex form (table 7).

Before specifying this effect of age, it was necessary to determine what this shift to the Evaluate strategy meant for the application of the procedures. Since the Evaluate strategy includes the DESCRIBE procedures, a relevant question is whether the Evaluate strategy influences the execution of the DESCRIBE procedures. As the adults did not use the Describe strategy in its pure form, they were excluded from this comparison.

It appeared that the number of applications of the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE did not show a relation with one of the two strategies (see table 8); neither did the discontinuous application of CHOOSE THEME, nor the size of the discontinuities (see table 9 and 10). This means that the Evaluate strategy did not affect the execution of the embedded DESCRIBE procedures. Thus, in order to examine the effect of age in more detail, the strategies could be lumped together.

A strong effect of age showed up between the age groups 10, 12, and 15 in the *frequency* of applying the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE (table 11). The older writers applied these procedures more often, presumably because they have a greater knowledge base to retrieve information from.

However, the *order* in which they retrieved information did not differ much from the younger ones. Older writers did not characterize the topic in a thematically more coherent way than the younger ones (table 12).

The increase of presented information with age (table 10) was reflected in the sizes of the first and second mentionings; both increased significantly. However, the relative size of the reintroductions did not differ much; the distribution of the information over the first and second mentionings was almost equal for the 10-, 12- and 15-year-olds. This means that the development of the 15-year-old writers manifests itself in generating more information in the first and second discussion of a theme, not in a more thematical coherent presentation.
The further development from 15 year of age on to adulthood showed an effect with age as well. However, this effect did not consist of an ongoing increase in the number and size of applied procedures. Adults applied the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE far less often than the 15-year-olds (table 13). It might be the case that those writers are more concerned about the relevance of the information, in contrast to the younger ones who mainly seem to be concerned about the completeness of the information (cf. Bereiter & Scardamalia 1987: 185).

The adults concentrated their description in first mentionings of themes. This appeared most clearly from a decrease in the (relative) number of discontinuities, and a decrease in the (relative) size of the second mentionings (table 14). These results suggest that adults reflect on the order in which information is retrieved.

In addition, a slight increase was found in the number of added evaluations in the introduction of the topic. Adults more often remarked at the beginning of their texts that they would not like to be like someone else. However, they also explicitly stated their cooperative attitude towards the assignment. With respect to the evaluations in the Description section and Conclusion of the text, the adults did not behave very differently from the younger writers.

b. Do the strategies correlate with the choice of the topic?

In the preliminary exploration of the corpus reported in chapter 3, it was decided to control the corpus for subject features such as age, type of school, and sex, as well as for the text feature 'interpretation of the assignment' (see chapter 3, section 3.5.1). This showed that most subjects had interpreted the assignment as 'describe an existing person', an exception being the Grammar school pupils, who more often constructed a person.

We decided to examine the texts describing a constructed topic independently from the relation with age. These texts, written by the 15-year-old Grammar school pupils, were used tc examine the effect of the choice of the topic (constructed versus concrete). Therefore they were compared to the texts of the peers describing a concrete person (15-year-old ULO-pupils).

The results showed that this distinction in choice of topic was useful as there was a clear relation between the use of strategy and the choice of topic. Writers describing a constructed person more often applied the Describe strategy, than those describing a concrete person (table 16). In addition, in applying the procedures CHOOSE THEME and ADD CONTEXT/ELABORATE, the writers of a constructed topic behaved differently as well. In both strategies, they applied the procedures far less often than their peers describing a concrete topic. This may be due to the fact that describing a constructed person requires more mental effort than just describing a concrete person. Constructing a person requires combining characteristics from existing persons and making up traits if the existing persons do not have the desired traits. Thus, describing a constructed person is more than activating one mental representation of a person and retrieving information (cf. Linde & Labov 1975). Several mental representations of concrete persons have to be activated. Subsequently, the most desired characteristics have to be extracted and assembled in a coherent way.

It appeared that the Grammar school pupils succeeded in presenting the assembled characteristics in a coherent way, as their (relative) number of discontinuities was much smaller than that of the ULO-pupils. Consequently, the (relative) size of the second mentionings was far smaller.

Finally, it appeared that there were only small differences in the number of explicitly stated evaluations of both types of topic. This might indicate that the evaluation of a constructed topic had already taken place in constructing the person. Negative traits that failed in the test of appropriateness were excluded from the description and could therefore not be evaluated within the text. Thus, the writers of the constructed topic did evaluate the topic and task, but this evaluation is not of a text-internal kind but of a topic-internal kind; the evaluations are incorporated in assembling the topic.

6.5.2 Evaluation of TRACE as indicator for development

Before framing the results, the instrumental strength of TRACE to reveal developmental changes will be evaluated.

With respect to TRACE as *instrument* to reveal developmental aspects, the following remark should be made.

TRACE is a Text-based reconstruction of conceptual activities. This means that it only captures conceptual operations as far as they leave traces in the final text. This restriction was already discussed in chapter 5. However, there is also a consequence for studying development. TRACE does not reveal operations that are not textually manifested. This does not seem to be a problem for young writers. The differences between what they think and what they write down are very small (see, for instance, Bereiter & Scardamalia 1987: 19). Besides, as was stated in chapter 3, younger writers are not very capable in expressing meta-linguistic operations. To study their processes TRACE seems to be an adequate instrument, as it does not ask for their meta-linguistic skills.

The writing skills of more skilled writers, however, might be underestimated. The expectation is that older writers, such as the adults and Grammar school pupils in our corpus, execute more operations than are retraceable from their texts; presumably, they think more than they write. Their (Reflect-)operations are probably more complex than TRACE can reveal. Thus, in studying expert's writing behaviour other methods, for instance, pause-registration, thinking-aloud, or interviewing, would serve as an adequate supplement. The argument presented in chapter 3 that these methods presuppose meta-linguistic knowledge and the skills to express them holds less for these older writers. Especially these older writers are assumed to be capable of thinking and expressing themselves on a meta-level about their writing process.

With respect to the *results* TRACE yields, we should consider what these developmental aspects contribute to the theoretical characterization of novices and experts. In our opinion, they are a specification of the general characterization discussed in chapter 2.

The main result that older writers, from the age of 15 on, use the Evaluate-strategy more often is congruent with the developmental scenario proposed by Bereiter and Scardamalia (1987). This change from knowledge-telling to knowledge-transforming requires problem solving effort, internalization of the feedback loop, and self-reports of cognitive changes (Bereiter & Scardamalia 1987: 246-247). Evaluations may be regarded as instantiations of (low level) problem solving. This developmental scenario describes the increasing gradual differentiation of Reflect from Select which leads to the use of different strategies.[10]
In novice writing, Reflect and Select operations are highly connected.[11] The leading strategy therefore stays close to the procedures executed in Select: it is a small reformulation. In expert writing, on the other hand, Reflect is differentiated from Select, which appears from the operations executed. Experts do not only describe the person, but also evaluate the person described, given the presented increase from the Describe to Evaluate strategy (table 7).

Table 5. The total number of texts in relation to age, type of school, and sex

age	type of school	girls	boys	total
10	primary school	49	35	84
12	primary school	94	60	154
15	secondary (LBO)	56	49	105
15	secondary (ULO)	47	45	92
15/16	secondary (Grammar school)	44	60	104
Total		290	249	539

b) Text features

The texts were also controlled for interpretation of the assignment. The subjects were asked to describe a person they would like to resemble. A number of subjects (N=88) did not comply with this assignment. Some of them did not produce any text at all, or wrote a text shorter than three lines. Others gave an enumeration of characteristics without formulating complete clauses, or wrote a narrative without describing a person. There were also subjects who wrote about themselves, and other subjects who described several persons instead of one. These texts were excluded from the sample and are referred to as 'drop-outs'. The number of excluded texts per group is presented in table 6 in absolute numbers and in proportions.

Table 6. Number of 'drop-out' texts in relation to age, type of school and sex (proportion between brackets)

age	type of school	girls	boys	total
10	primary	6 (.12)	2 (.06)	8 (.10)
12	primary	16 (.17)	8 (.13)	24 (.16)
15	secondary (LBO)	7 (.13)	6 (.12)	13 (.12)
15	secondary (ULO)	9 (.19)	7 (.15)	16 (.17)
15/16	secondary (Grammar school)	9 (.20)	18 (.30)	27 (.26)
Total		47 (.16)	41 (.16)	88 (.16)

The proportion of drop-outs did not differ for the sexes (both 16%). The proportion of drop-outs per group is statistically equal for all groups.[19] All these drop-out texts (N=88) were excluded from further exploration.

Although the remaining 451 subjects complied with the assignment, a closer examination showed that not all of them had interpreted the assignment in the same way. The following five interpretations could be distinguished:

a) describe an existing person;
b) describe an existing person and attribute ideal aspects to him;
c) describe a constructed person and illustrate the description with one or more characteristics of an existing person;
d) describe a constructed person on the basis of characteristics of different existing persons[20];
e) describe a constructed person without reference to existing persons.

Table 7 presents the number of different interpretations of the assignment (a to e) per group in absolute numbers and proportions. Because the interpretations b, c, and d occurred infrequently, they were lumped together.

Table 7. Interpretations of the assignment in relation to age, type of school and sex

		girls				boys			
age	type of school	a	bcd	e	total	a	bcd	e	total
10	primary	.91	.05	.05	43	.91	.03	.06	33
12	primary	.78	.08	.14	78	.98	.00	.02	52
15	secondary (LBO)	.79	.10	.10	49	.84	.05	.12	43
15	secondary (ULO)	.74	.10	.16	38	.89	.05	.05	38
15/16	secondary (Grammar school)	.40	.20	.40	35	.40	.12	.48	42
Total		.74	.10	.16	243	.79	.05	.16	208

The majority of the subjects interpreted the assignment according to the a-version 'describe a concrete person'. However, this preference decreased with age from over 90% to about 40% (girls: $\chi^2(4)=29.05$, $p<.001$; boys: $\chi^2(4)=58.75$, $p<.001$). The major jump is found between 15-year-old secondary (ULO) pupils and those of the Grammar School.[21]

Closer examination of the texts in the a-category showed that two kinds of concrete persons were described:

- acquaintances from the private neighbourhood of the subjects, for example a mother, a brother, a schoolfriend or a teacher;
- public figures with a(n) (inter)national reputation, such as a politician, a popstar or a sportsman.

The 15-year-olds are in the middle of the scale; for some operations, they resemble novices (*Select-organize* mentioned in table 20; cf. results in table 18); for others, they behave more like adults (*Select-retrieve*; cf. results in table 17).[15]

A tentative explanation of these differences would be based on general cognitive developmental changes, such as an increase of knowledge and skill, which of course underlie this shift. However, there might be a more specific explanation. The shift from describing to evaluating, from Select to Reflect, might be ascribed to a different way of writing for which we would like to quote Bereiter and Scardamalia (1987: 83-84). Reflect-operations can only occur in an "iterative language system". The Reflect-operations require an executive mechanism for switching between a forward process of text generation (guide-operations) and a backward process of evaluation. If children, switch from generating information to evaluating it or vice versa, they have the problem of switching back, because they have lost track of where they were in the writing process. This illustrates that Reflect-operations require extra capacity.

Notes

1. Its corresponding PISA structure is presented in the Appendix to chapter 6 in figure A.
2. See Appendix to chapter 6 table A, for the original Dutch version of the text.
3. There is only one simpler way which is just mentioning the topic. However, these kinds of 'texts' are excluded from the sample as was discussed in chapter 3, section 3.5.
4. See Appendix to chapter 6 table B, for the original Dutch version of the text.
5. See Appendix to chapter 6 table C, for the original Dutch version of the text.
6. The descriptions about acquaintances and public figures are lumped together, since there was only little difference between those descriptions (this distinction was made in chapter 3). Only with respect to the size of two themes a small difference was found. The descriptions of acquaintances contained slightly more information related to the theme Hobbies; descriptions of public figures contained slightly more information about the theme Career.
7. One would expect that a writer who uses the Evaluate-strategy and chooses to describe a theme for the second time, will mark this second mentioning. Using these pop-markers is, however, hardly the case. Probably this is caused by the fact that such a second mentioning is used to cover a different range of information belonging to the theme.

This result elicitates a critical remark about the thematic categorization of the text; it might be the case that the themes used as cues in the assignment do not completely correspond to those presented in the human mind, indicating a description. This, however, is an empirical question which is interesting to answer in further research.

8. A more specific about the influence of Evaluate could be that the distance between the first and second mentioning is smaller and that an explicit pop-marker is used under the influence of Evaluate.

9. Only in a few cases does a writer reopen a theme for the third or fourth time. The sizes of these reopenings are included in those of the second mentionings.

10. Bereiter and Scardamalia distinguish between rhetorical strategies and self-regulatory strategies. The rhetorical strategies are used for achieving effects on purposes of the composition. Those are not the issue in either their study or in this one. The strategies that are referred to are self-regulatory strategies, used to manage one's own behaviour (Bereiter & Scardamalia 1987: 249-250).
A related issue is that strategies need something to work on, that is mental representations (1987: 352). What remains unknown is whether young writers do not know the (right) strategies, or do know them, but do not master them in the right way, for instance because the mental representation(s) are ill-defined (for the interdependence of strategies and mental representations see Bereiter & Scardamalia 1987: 325 ff.)

11. In addition it should be remarked that the Reflect-operations executed by novices differ from those of experts. We will leave aside the discussion about to what extent novices's operations differ (do they execute the same operations but in a less profound way, or do they operate completely different).

12. Strictly speaking TRACE does not reveal the intentions of a writer. However, on the basis of the theoretical insights that novices have a task executing way of writing and the experts a problem solving way (Flower & Hayes 1980), it is plausible to assume that the presented intentions, which are very low-profile defined, play a role in describing a person.

13. Bereiter & Scardamalia have devoted a complete chapter to 'coordinating ideas' (1987: chapter 6).

14. It should be remarked that the performance of the adults is perhaps influenced by the fact that they had to describe a concrete person. If they would have been free to chose a topic they probably would have chosen a constructed person, given their evaluative remarks in the introduction.

15. Bereiter and Scardamalia remark that as soon as the writer aspires to something beyond minimally adequate text, however, the processing demands of composition can mount rapidly. These additional processing demands may arise from the addition of rhetorical constraints (..) or from demands for deeper processing of content (1987: 152-153)

7 Conclusions

7.1 Summary of the study

The blueprint of a writer presented in *chapter 2* is a psycholinguistic model based on process theories of writing (Bereiter & Scardamalia 1987; Flower & Hayes 1980, 1981) and of speaking (especially Levelt 1989; Dijkstra & Kempen 1993). It describes the conceptual processes in terms of Reflect, Select and Shape operations. In *Reflect*, a writer considers intentions, strategies, and expectations that guide Select and Shape operations; these intentions and strategies are also used to control Select and Shape operations. Reflect checks for agreement between the Selected and Shaped information on the one hand and the intentions and expectations on the other. In *Select*, information is retrieved on the basis of cues, and then tested for its appropriateness. In addition, local organization patterns are determined. Before this selected information can be expressed verbally, a suitable presentation form for the information units is considered. *Shape* decides on aspects such as the status of the information and the topicalization of referents. This study is restricted to the fundamental activities Reflect and Select, since they are regarded as the basic operations within the conceptual component.

The above picture of conceptualizing is obviously the expert's way of writing; novices operate differently. In line with Bereiter and Scardamalia's developmental scenario, novices do not (or only occasionally) execute Reflect operations. While they do so later in their development, their way of Reflecting is strongly related to the Select operations. The development to expert writing starts when the Reflect operations are detached from the other operations.

The nature and development of these conceptual operations can be characterized by the strategies and procedures applied: experts apply other strategies and procedures than novices. However, - to the best of our knowledge - it is not known *exactly* what strategies and procedures writers do apply. In this study we have tried to contribute to an *explicit* and *systematic specification* of the nature and development of the conceptual operations. Using an empirical approach, we have attempted to gain insight into these operations by reconstructing them.[1]

In order to specify the conceptual operations, a methodological issue had to be dealt with first, i.e., 'How can Reflect and Select operations be studied?'.

There are on-line and off-line methods which are either intrinsic or extrinsic to the writing process. The extrinsic methods (such as thinking aloud and interviewing) require meta-linguistic skills of the subject. These methods are therefore not suitable for subjects of all age levels. Young writers in particular lack the required meta-linguistic skills (Bereiter & Scardamalia 1987: 70). As this study examined the conceptual processes in writing from a developmental point of view, it was necessary to include young writers as well. In addition, we wanted to study 'natural' text production, with as little interference from the researcher as possible. Therefore, it was decided to use an off-line method: *text analysis*.

This, however, implied the presence of adequate textual manifestations. The texts used in this study originated from a corpus of descriptive texts written by children varying in age from 10 to 15/16 years (N=74). In order to study the development from the age of 15 to adulthood, texts written by adults were added to this corpus.

Before taking the sample, the original corpus was explored for subject features (age, type of school, sex) and text features (realization of the assignment or not, type of topic described) (see *chapter 3*). This exploration showed that it is necessary to control the sample not only for subject features but also for text features in order to have a homogeneous and representative sample of texts (see Van Wijk 1987; Van Wijk & Luiten 1987). The sample used in this study was checked on the classic developmental indices on the sentence level. This showed that the sample was not atypical; the T-unit length and subordination-index increased regularly with age and was of the same range as obtained in other studies (cf. Hunt 1965; Loban 1976; Van Wijk 1987).

Text analysis as a means of providing insight into conceptual processes further implies that it is known what clues in a text reveal (parts) of these processes. Empirical studies indicate that the structure of a text in combination with the content mirror the Reflect and Select operations best (Bereiter & Scardamalia 1987: 206). That was the main reason for analyzing the structure of the texts (in combination with the content).

If text analysis is used as an instrument to reduce texts to comparable structures, the instrument has to be explicitly formulated in order to yield reliable results. It is necessary to prescribe what linguistic elements on what level have to be *inspected* and what relevance they have with respect to the *connection* of segments. In addition, to achieve a coherent *integration*, the global relations between text parts have to be taken into account. Existing methods, however, are not yet sufficiently explicitly defined; they do not specify in detail how linguistic features indicate text structure and they leave much to the intuition of the analyst. Nevertheless, these methods provide a fruitful basis for developing the procedural method, called *PISA*, an acronym for Procedures for Incremental Structure Analysis. In PISA, ideas from psycholinguistic and linguistic theories have been integrated. The basic principles of the

first version are described in Sanders (1992), and Sanders and Van Wijk (1995). The further development of this instrument has led to a number of improvements, consisting of extensions and a reorganisation, which resulted in a second version of PISA described in *chapter 4* (see appendix to chapter 4 for a detailed presentation).

PISA assigns a labelled hierarchy, consisting of a *microstructure* in which segments are coordinated or subordinated, a *macrostructure* in which groups of segments are related to themes within the topic of the text, a *superstructure* in which the function of segments within the text is indicated, and a labelling of the *rhetorical relations* within these structures. This output results from an INSPECTion of linguistic elements present in the segments; the CONNECTion of the segments in the microstructure takes place on the basis of a set of condition-action pairs. The INTEGRATion to the superstructure takes place on the basis of explicit instructions for the analyst and by making use of a knowledge base with encyclopedic knowledge.

If we apply PISA to the examples presented in (1) and (2), which were also used as illustrations in chapters 1 and 4, PISA assigns the structures presented in figure 1 and 2.[2]

(1)
1 He is a teacher in Heervarenbeek.
2 He teaches children to learn
2a like arithmetic Language History Geogaphy.
3 He is round about 40 years old.
4 He wears a suit and a white shirt, and black shoes and looks healthy.
5 Children teaching.
6 On Sundays, Wednesday Afternoons, and Saturday Afternoons cycling
6a Correcting Notebooks
6b Walking in the woods.
7 His naame is Uncle Harie.
8 Zometimes he goes to tilburg To the families
8a and window-shopping on Sundays as well.

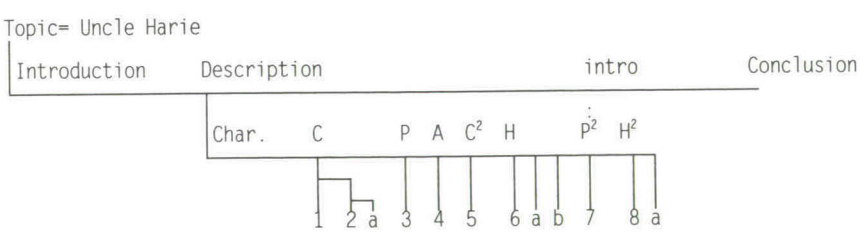


Figure 1. Structure graph of text 1 presented in (1)

(2)

1	I would,
1a	if forced to make a choice
1	like to be like Ed Nijpels.
2	He looks quite nice,
3	is well-dressed,
4	and expresses himself very well.
5	Besides, he has undertaken all kinds of activities alongside his rather varied career.
6	It is this variation that appeals to me very much.
7	His "main job" at this moment is being mayor of Breda,
8	but as a compere on TV he holds his own as well!
9a	When he relaxes after his busy work,
9	he can be found on the tennis court
10	and the evenings he likes to be with friends or just on his own reading a nice book!
11	Because of his open character he makes friends easily
12	and makes contacts easily.
13	As a mayor he tries to stand close to the people.
14	The age of Ed Nijpels
14a	(round about 40)
14	is an age which has many advantages:
15	one has had the chance (and the time!) to make something of one's life.
16	one does not need to be a 'he-man' any more.
17	Summarizing, I must say that in any case it looks as if Ed Nijpels feels comfortable
18	or
18a	to put it informally
18	'he likes who he is'
19	and thàt is something worth striving for!

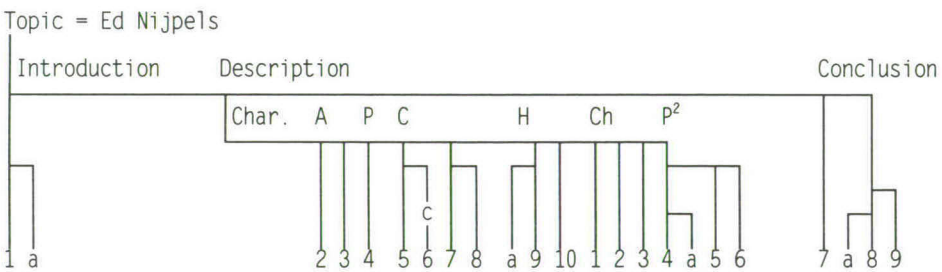


Figure 2. Structure graph of text 2 presented in (2)

These structures provided the basic data for our study; they were interpreted in close connection with the specific contents of the segments (cf. Van Wijk 1992).

With regard to the *superstructure* of the text, the analysis shows that the second text consists of three parts: an Introduction of the topic, a Description section consisting of a list of characteristics, and a Conclusion. In the first text, the introduction and the conclusions are missing. The text starts without introducing Uncle Harie. His name is not mentioned until

segment 7. The writer does not round off his text very consciously either; he just stops writing. There are no conclusions or summary statements like those in the second text.

The *macrostructure* shows the elaboration of the Description section; it shows *what* themes are discussed and *how* they are elaborated (the themes are abbreviated on the Characteristics line). These aspects indicate what information is presented. In the first structure, the description covers four themes: nothing is said about the character of Uncle Harie. The structure of the second text shows that the description covers all five themes. Another structural aspect concerns the *thematical coherence* of the information presented. The macrostructure shows whether all the information about one theme is grouped together, or whether the same theme is discussed in different parts of the text. In other words, it reveals thematical discontinuities in the structure. The discontinuities are indicated by a superscript at the abbreviation on the characteristic line. If we look at the Description-section of the structure in figure 2, we see a fairly well-ordered presentation. Only the 'personal information' is not presented as one cluster. In contrast, the text in figure 1 has three discontinuities in discussing the themes. In segments 6, 8 and 8a, the activities of the person described during his spare time are described. But this description is interrupted in segment 7 by the presentation of the name of the topic; segment 5 is a repetition of segment 2, which also indicates a disorderly presentation.

The *microstructure* provides insight into the stratification of the text. The structure of the first text is rather shallow; the description mainly consists of a list of characteristics. The structure of the second text, on the other hand, is more stratified. Although this text is not as stratified as several others in the corpus, it illustrates the idea of an hierarchy. The text in (2) consists of an enumeration of characteristics as well as subordinated segments which specify previous segments, such as segments 14a, 15, and 16, specifying segment 14. In addition to this, contrast relations are explicitly stated, as in segments 7 and 8, and arguments, as in 1a to 1 are given.

If we assume that writing processes leave traces in the text (see, for instance, Cooper 1983; Tamor & Bond 1983; Bereiter & Scardamalia 1987: 41), the PISA structures can be regarded as the results of (final) conceptual activities. On the basis of this assumption, the structures were used to reconstruct parts of Reflect and Select operations. The reconstruction resulted in a TRACE model, in which *TRACE* is an acronym for Text-based Reconstructions of Activities in the ConceptualizEr. The model represents conceptual operations in terms of a set of reconstruction rules. The set consists of two main rules, one corresponding to Reflect operations and the other to Select operations. These main rules include more specific rules.

In *chapter 5*, a TRACE model for writing a description of a person is proposed. This TRACE model consists of a set of rules corresponding to procedures executed in writing a text. The rules are called by two main procedures: the DESCRIBE procedure and the EVALUATE procedure which governs the DESCRIBE procedures. The set of DESCRIBE procedures corresponds to Select operations; the set of EVALUATE procedures corresponds to Reflect operations.

If we apply TRACE to the two examples, it yields the reconstructions presented in tables 1 and 2.

Table 1. Reconstruction of conceptual activities in writing the text presented in (1), on the basis of the PISA structure presented in figure 1. The output denotes the segments of the text

TRACE-rule		Segment
DESCRIBE Topic		
→ **INTRODUCE Topic + CHARACTERIZE Topic**		
CHARACTERIZE Topic*		
→ CHOOSE THEME	→ Career	1
→ ELABORATE THEME	→ Specify	2
→ ELABORATE THEME	→ Specify	2a
→ CHOOSE THEME	→ Personal Particulars	3
→ CHOOSE THEME	→ Appearance	4
→ CHOOSE THEME	→ Career	5
→ CHOOSE THEME	→ Hobby	6
→ CHOOSE THEME	→ Hobby	6a
→ CHOOSE THEME	→ Hobby	6b
INTRODUCE Topic	→ Name	7
CHARACTERIZE Topic*		
→ CHOOSE THEME	→ Hobby	8
→ CHOOSE THEME	→ Hobby	8a

Key: '*' means recursive use

The reconstruction presented in table 1 shows that this writer used a restricted set of procedures, the DESCRIBE-procedures only (INTRODUCE Topic + CHARACTERIZE Topic), in generating his text. In using these procedures, he is also rather limited. Again and again he applies the same procedure (CHOOSE THEME) to characterize the topic; he only elaborates a presented aspect twice (segment 2 and 2a). In choosing a theme, the writer also reintroduces themes already discussed (Career and Hobby).

The order in which the writer applies the main procedures (INTRODUCE Topic + CHARACTERIZE Topic) is remarkable. He does not start by introducing the topic, but by

presenting characteristics.[3] After having presented several characteristics, he mentions the name of the topic; subsequently, he continues characterizing Uncle Harie.

Table 2. Reconstruction of conceptual activities in writing the text presented in (2), on the basis of the PISA structure presented in figure 2. The output denotes the segments of the text

TRACE-rule			Segment
EVALUATE (DESCRIBE Topic)			
→ **EVALUATE (INTRODUCE Topic)**			
+ **EVALUATE (CHARACTERIZE Topic)**			
+ **EVALUATE Description**			
EVALUATE (INTRODUCE Topic)			
→ INTRODUCE Topic	→	Mention name	1
→ EVALUATE (INTRODUCE)	→	Negative attitude	1a
EVALUATE (CHARACTERIZE Topic)*			
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→	Appearance	2
→ CHOOSE THEME	→	Appearance	3
→ CHOOSE THEME	→	Personal particulars	4
→ CHOOSE THEME	→	Career	5
→ SPECIFIC-EVALUATION	→	Positive opinion	6
→ CHARACTERIZE Topic			
→ CHOOSE THEME	→	Career	7
→ ELABORATE THEME	→	Specify	8
→ ADD CONTEXT	→	Condition	9a
→ CHOOSE THEME	→	Hobby	9
→ CHOOSE THEME	→	Hobby	10
→ CHOOSE THEME	→	Character	11
→ CHOOSE THEME	→	Character	12
→ CHOOSE THEME	→	Character	13
→ CHOOSE THEME	→	Personal particulars	14
→ ELABORATE THEME	→	Specify	14a
→ ELABORATE THEME	→	Specify	15
→ ELABORATE THEME	→	Specify	16
EVALUATE (Topic and Description)*			
→ **EVALUATE (Topic)**			
→ Summarize			17
→ Summarize			18a
→ Summarize			18
→ **EVALUATE (Description)**			19

The reconstruction presented in table 2 shows that the second writer used a more extended set of procedures than the first writer (see table 1). In addition to the DESCRIBE procedures, the writer has also applied the EVALUATE-procedures. He introduces the topic and adds a negative

evaluation of the assignment (segment 1a). In the Description section, the writer also makes an evaluative remark with respect to a presented aspect (segment 6). In using the DESCRIBE procedures, this writer elaborates themes discussed (segments 8, 14a-16); in one case he adds a context to a characteristic (segment 9a). In contrast to the first writer, this writer rounds off the description with a summary statement. He concludes with the general impression he has of Ed Nijpels (18) and uses it to indicate the most salient characteristic which makes him want to be like Ed Nijpels.[4]

These individual reconstructions were used to gain insight into the development of the processes with age. TRACE rules reveal how writers have presumably operated, in other words, the TRACE rules can be regarded as the *procedures* executed by the writer. In the case of the examples, the TRACE tables show that the first writer applied only the DESCRIBE procedures, whereas the second writer used EVALUATE procedures as well.

The procedures were used to characterize individual performances. Subsequently, comparison of these characterizations revealed regular patterns of applying the procedures for groups of writers, directed by *strategies*. These strategies played a crucial role in modelling the development of the conceptual processes. On the basis of the TRACE rules for describing a person, two main strategies have been postulated: a Describe strategy, consisting of DESCRIBE procedures, and an Evaluate strategy, consisting of EVALUATE procedures. The writer of the first text, the 10-year-old boy, exemplifies the use of the Describe strategy; the writer of the second text, the adult, the use of the Evaluate strategy. This relation between age and strategy corresponds to the results of the entire corpus presented in *chapter 6*. When writers get older, they use the Evaluate strategy more often. This fits in with what Flower and Hayes state: 'good' writers solve different problems (1980: 30)).

7.2 The scope and limitations of this study

This study has resulted in a proposal for an *explicit* and *systematic specification* of the procedures and strategies involved in the conceptual processes in writing, i.e., a TRACE model for descriptive texts. It is a contribution to psycholinguistic research on language production on the *text* level, in an *empirical* approach.[5]

TRACE

From a theoretical point of view, the TRACE model contributes to a specification of the macroplanning in Levelt's blueprint for a speaker. As Levelt remarks, the conceptualizer and the message generator can be thought of as a structured system of procedures (Levelt 1989: 10). In Levelt's blueprint for a speaker, however, the emphasis is on microplanning; macroplanning has been given less attention. TRACE provides an initiative to specify these conceptual processes for a specific domain, i.e., writing descriptive texts.

With respect to the models developed by Bereiter and Scardamalia, and Flower and Hayes, the present study is distinguished by the use of the empirical and text analytical approach. Bereiter and Scardamalia point out that "we need to figure out the rules that novices use in composing (...) and that the most direct way to discover such rules is through analyzing texts" (Bereiter & Scardamalia 1983: 23). They emphasize the use of text analysis - thus using empirical data - as a research method. They argue that probably "nothing is holding back progress toward understanding the composing process so much". Bereiter and Scardamalia subsequently indicate more precisely what they think a discourse analyst should do to contribute to the rule system, that is, developing a 'child rhetoric'. They state that it should not be simply "rating children's texts for this or that characteristic or enumerating the ways in which they differ from texts of mature writers; rather, it means working out a rule system that gives rise to texts like those children write" (1983: 23). This quote indicates that text analysis is only an *instrument* in this cognitive research, it does not yield the final results. In our case, PISA is the instrument for describing *texts*, not processes. The structures it yields are used to work out the rule system, TRACE, which indicates what *processes* have presumably been executed. This was done for a writing assignment that left many of decisions about content and organization to the writer (see Levelt 1982 and Linde & Labov 1975 for a rule system for a conceptually prestructured task).

The specificity of the rule system proposed in this study was not obtained at the cost of the applicability of the rules to the subjects' texts. The TRACE rules seemed to yield a successful reconstruction for all texts and made it possible to characterize the developmental level (this in contrast to McCutchen & Perfetti 1982 who left out 60% of their sample in order to arrive at their purported characterization; or Witte & Cherry 1986 who propose a rather general system).

Moreover, this specificity has the advantage that it makes it possible to use these results in the field of computational linguistics for text generation (cf. Paris & McKeown 1987[6]).

One restriction of TRACE is that it does not provide insight into why a writer operates the way he does. Nevertheless, with a little bit of creative thinking one can guess why a writer makes, for instance, certain theme switches. See, for instance, the two text examples (1) and (2).

The theme switches in the description of Ed Nijpels are almost 'natural'; they seem to be directed by the previous segments. Take, for instance, the switch from Career to Hobby, from segments 8 to 9a. This switch is a rather smooth one because the author refers explicitly to the temporal order of activities: first, he is busy working, *after that* he relaxes'.

Not only can such temporal aspects establish coherent switches, semantic overlap may have that effect as well. Take the switch from Hobby to Character, from segment 10 to 11. The mentioning of 'friends' in both segments makes the switch less disconnected. Although 'friends' in segment 10 refers to specific persons, and in segment 11 to a generic group, there is semantic overlap which creates content relatedness. The transition from segment 12 to 13 can be characterized in the same way. 'Making contacts easily', on the one hand, and 'standing close to the population', on the other, both relate to social aspects. The first describes Ed Nijpels' character and the second its repercussion in his career (which causes the switches in themes).

It seems that the theme switches in the Ed Nijpels text -in contrast to the ones in the text about Uncle Harie- are triggered by information in previous segments. This might indicate that the way adults retrieve information is different from the way children do. We suppose that adults more often use *text-internal* cues -the text produced so far- for information retrieval, whereas the younger writers probably need *text-external* cues, such as the assignment, for information retrieval.

A critical remark should be made with respect to disregarding Shape aspects. The reconstructions are based on structure analyses in combination with the original text, which results in less attention being paid to the Shape-aspects.[7] Nevertheless, Shape has an indirect influence on the reconstructions. The PISA structures are based on linguistic elements present in the text; these elements and their position, obviously, result (partly) from Shape operations. Presenting a characteristic, for instance, has repercussions for referring to that theme; if it is a new theme, a full noun phrase is more suitable than a demonstrative.

The interaction and mutual dependency between the different conceptual processes is, however, (still) hard to describe. This problem is related to more general discussions on cognition about whether the processes are modular or not, what their exact input and output is, etc., which make it a central problem in psycholinguistics in general.

PISA

In addition to TRACE, this study resulted in the further development of the analytical instrument PISA. PISA was developed for and is used in studying conceptual processes involved in writing expository texts in general, which means that the scope of PISA goes beyond what is needed for this study.

How can this extended version of PISA be evaluated?

First of all, more *theoretical* ideas about how people write and read texts have been incorporated in the extended version. In addition to a more sophisticated micro-structure, the inclusion of a super- and a macrostructure provides a more solid basis for inferring processes from the products.

Second, from an *instrumental* point of view, the extended version of PISA can also be evaluated positively. PISA makes it possible to compare texts written by subjects differing in age and skill. This is an important condition in studying developmental changes. The structure assigned by the extended version of PISA revealed salient differences especially on the macro- and superstructure level; thus, the extensions were crucial.

Finally, from a *technical* point of view, it should be noted that it is hard to define strict rules for connecting segments to the macro- and superstructure, which restricts the possibilities for implementation. What we succeeded in is making explicit instructions for the analyst. Another consequence of the extensions is the use and size of the knowledge base. In an analysis of descriptions of persons, the knowledge can be confined to a list of items specifying the themes of the macrostructure and linguistic elements indicating a relation to the superstructure (such as 'Summarizing' in segment 17 of text 2) However, in analyzing, for instance, juridical texts, the use of knowledge of the world has to be much more extended (see Schilperoord 1995).

Educational implications

Research and education of writing skills focus mainly on guiding Reflect operations (cf. planning). However, good writers not only distinguish themselves by more and different planning, they also evaluate (cf. revise) their texts more often and on a different level; revision takes place not only on the word and sentence level but also at the text level. From a developmental and educational point of view, it is therefore not only interesting to know how the developmental stages can be characterized in terms of strategies and procedures (how the Reflect and Select develop with age), it is also useful to know how the texts of 'poor' writers can be revised in order to improve their quality.[8] As Bartlett (1982: 354) states with respect to the importance of Reflect operations in an educational context, 'the skill of revision is

related to skill of formulating problems and that again is related to formulating plans and goals'. This conforms to the operations depicted in the blueprint of a writer;[9] by adding Reflect operations, writing becomes a problem-solving activity.

In order to have (moderate) writers perform better -and influence their competence as well- more attention to revision is essential in writing education. Revision must not be restricted to encoding activities; it has to focus on the retrieval and organization of information (Select). The idea is that concentrating on Reflect and Select operations in revision will improve the quality of the text.[10]

From a developmental point of view, we are interested in the improvement of texts with respect to the level of writing skill (Van Wijk 1992, Van der Pool 1993a, 1995a,b,c). In improving text quality, we take the level of proficiency into account; we do not use texts of educated writers as a model for revising texts produced by immature writers. Assuming that adapting a written text is easier than planning a text which still has to be written, we emphasized the importance of the revision-part in the writing process. Our revision addresses a reordering of segments, as this affects more than just the coherence of the presented information.[11]

The structures PISA assigns to a text provide an opportunity to offer revision advice to students. With marginal organizational changes, texts can be improved, simply by reordering segments on a global level. This can be illustrated by the first text about uncle Harie.

With regard to the superstructure of the text, the analysis showed that the introduction of uncle Harie interrupts the Description section and that a conclusion is lacking. As the introduction and conclusion of a text have a functional role for the reader, they also improve text quality.[12] In addition, the writer reintroduced the same theme at different moments in the text, which resulted in a thematically incoherent description. The order of the information is crucial for the reader in building a (coherent) representation of the text. Therefore it plays an important part in the comprehensibility of the text and is related to a criterion for text quality.

How can this structure provide a helping hand in revising the text?

The disorganization of the information signalled in the structure can be straightened out by moving segments and deleting a redundant one. The reordering consists of placing segment 7 at the beginning of the text, followed by segment 3. As segment 5 is a repetition of what was already discussed in 2, it can be deleted. These movements result in a thematically organized text; it yields the structure presented in figure 3, and the text in (3).

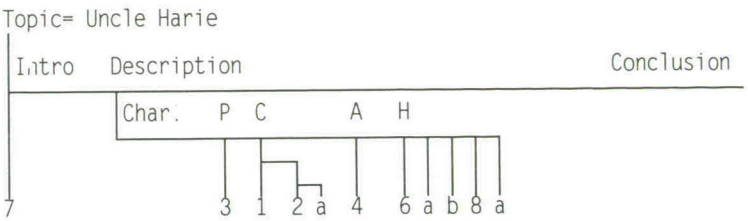


Figure 3. Structure graph of text 1 in revised form

(3)

7	His naame is Uncle Harie.
3	He is round about 40 years old.
1	He is a teacher in Heervarenbeek.
2	He teaches children to learn
2a	like arithmetic Language History Geogaphy.
4	He wears a suit and a white shirt, and black shoes and looks healthy.
6	On Sundays, Wednesday Afternoons, and Saturday Afternoons cycling
6a	Correcting Notebooks
6b	Walking in the woods.
8	Zometimes he goes to tilburg To the families
8a	and window-shopping on Sundays as well.

The revised text is obviously still far from perfect. The lack of a closing and a description that does not cover all themes means that the communicative function leaves much to be desired. The PISA analysis concerns only the organization of the text. It does not indicate how informative the text is with respect to the goal and readers: Does it offer sufficient information? Is all the information relevant? And is it correct?

However, in spite of these limitations, PISA may be a useful instrument for evaluating the effects on text quality with respect to the organization of information in an explicit and objective way. The structure analysis presented here is a step towards formulating criteria for text quality by taking psychological and text-linguistic considerations seriously into account.

7.3 Further research

Although this study provides answers to specific questions about human language production, it also raises questions about related aspects.

This study suggests the need for an extension with on-line evidence for TRACE operations. This could be realized by combining the off-line data (text structures) with on-line data (pauses). In Schilperoord and Van der Pool (1995), a study is presented related to the combining of these data and using them to model the cognitive processes. This study could be extended in further research by making a rule system for the occurrences of pauses (a pause-based reconstruction), parallel to TRACE, and combining these two rule systems in order to find correlations in the assumed conceptual processes (a proposal for such an integrated model is presented in Schilperoord & Van der Pool 1995).

It also raises questions related to developmental aspects. How does language production develop with respect to elderly people? The TRACE model and the PISA instrument provide tools to investigate the further development of language production. This kind of gerontological research is rather new (Köpper & Bosshardt 1994; Van Wijk & Lee 1995).

In addition, the instrument used in this study, PISA, and the model proposed, TRACE, need further development. In developing PISA, further attention can be paid to the applicability of PISA to other types of text, such as argumentative texts and the occurrence of discourse patterns such as responses, as well as implementation. The further development of the TRACE model can be realized by making it applicable for other types of text. Van Wijk (1995) has made a start by presenting a TRACE-model for argumentative texts.

As this study has indicated, writers do not produce incoherent sets of sentences: they write texts. Therefore we would like to emphasize that research on language behaviour should not pay attention only to production on the word and sentence level. It is desirable to combine the separate insights into parts of language production to increase insight into the production on a discourse level. Such an interdisciplinary approach to research may contribute to gradually unfolding parts of 'the black box' of the human mind.

Notes

1. In exploring the operations, this study focused on the writing behaviour of individual writers, in contrast to developmental research in educational contexts in which mostly groups of writers are compared.

2. PISA also assigns relational meanings to the connections. For a specification of the relational meanings of the two examples, see chapter 4, section 4.3.

3. Given the lack of other evaluative procedures, it seems unlikely that the late introduction of the topic is intentionally directed in order to achieve a rhetorical effect. In one text of the corpus, there was also such a late introduction of the topic. However, the reconstruction of this text revealed evaluations that indicated an intentionally guided way of writing. The writer violates this general order consciously in order to achieve a rhetorical effect, i.e., keeping the reader's attention by making him curious. Probably the writer intends the reader to wonder who the text is about. In order to get an answer, he has to read the entire text.

4. The relation between this remark (segment 19) and the evaluation of the introduction (segment 1a) is worth noting in this respect. At the beginning of his description the writer signalled implicitly that he would rather not resemble anyone else; nevertheless, he chose a topic. At the end of the description, it appears that he would like to be like someone who also likes who he is, i.e., not wanting to resemble someone else.
   The fact that the adult writer has chosen a concrete person instead of constructing one may be due to the assignment; adult writers were instructed to describe a concrete figure, in order to have a homogeneous text sample.

5. This attention to the word and sentence level is evidently dominating psycholinguistic research; see, for instance, the recent handbook of psycholinguistics (Gernsbacher 1994) which mirrors this attention for the word and sentence level.

6. The notation they use, in the form of an augmented transition network, can be regarded as a notation variant of the rewrite rules of TRACE.

7. In an attempt to gain insight into these Shape aspects the corpus used in this study was also analyzed for the given-new distribution of the information. (For a description of the method developed for this analysis, see Van der Pool and Van Wijk 1991.)
   Corresponding to the results of McCutchen and Perfetti 1982, it appeared that most segments were connected on the basis of references (the topic of the descriptions), which resulted in list-like structures. Older writers displayed more local connections, which showed a slight zig-zag pattern. These local connections were most often signalled with connectors. Especially the number of causal connectors increased slightly with age. The preliminary exploration of the sample on syntactic features, described in chapter 3, may also be interpreted in terms of Shape. The increase of bound subordinated clauses may indicate that older writers apply different microplanning and/or formulation operations. The variations between the performances were too small to base a revolutionary rule system on.

8. For a discussion of the role of teaching strategies and procedures, see Van der Pool & Van Wijk 1995.

9. This also matches the results with respect to developmental changes presented in chapter 6. Older writers evaluate their descriptions more extensively.

10. A good introduction to this type of education, guided by the computer, is the tested programme SPIRIT (Van der Geest, 1991). This programme, however, concentrates on planning and production and leaves the possibility to guide revision aside (Van der Pool 1993b). The advantage that pupils can get on-line

comment on (re)producing their texts on the sentence and text level is not completely optimized. Although on-line (structural) analysis is still in the future, more attention should be given to revision.

11. The reason why we limit the revision to a reordering of segments is that this can be done in a controlled way. Other revisional changes, such as adding information, lack such control. It is rather speculative and subjective to add, for example, segments in the way a 10-year-old writer would have. Therefore we have chosen a modest revision of this kind. As a result of the reordering of segments, redundant (i.e. repeated) information is deleted.

12. In an experiment in which the texts used in this corpus were judged by a panel of experienced judges, it appeared, among other things, that the number of reintroductions (computed by their geometric mean) of themes correlated significantly with a negative judgment.

References

Allport, A., MacKay, D., Prinz, W. & Scheerer E. (1987). *Language perception and production. Relationships between listening, speaking, reading, and writing*. London: Academic Press.

Anderson, R. & Pitchert, J. (1987). Recall of previous unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behaviour, 17*, 1-19.

Andringa, E. (1990). Verbal data on literary understanding: a proposal for protocol analysis on two levels. *Poetics, 19*, 231-257.

Appelt, E. (1985). *Planning English sentences*. Cambridge: Cambridge University Press.

Applebee, A. (1978). *The child's concept of story*. Chicago: University of Chicago Press.

Balota, D., Flores d'Arcais, G. & Rayner, K. (1990). *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.

Bartlett, E. (1982). Learning to revise: some component processes. In M. Nystrand (Ed.), *What writers know. The language, process, and structure of written discourse* (pp. 345-363). New York: Academic Press.

Bereiter, C. & Scardamalia, M. (1983). Levels of inquiry in writing research. In P. Mosenthal, L. Tamor & S. Walmsley (Eds.), *Research on writing; principles and methods* (pp. 3-25). New York: Longman.

Bereiter, C. & Scardamalia, M. (1982). From conversation to composition: the role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in Instructional Psychology, Vol. II* (pp. 1-64). Hillsdale, NJ: Erlbaum.

Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.

Bergh, H. van den & Meuffels, B. (1993). Schrijfvaardigheid. In A. Braet & J. van de Gein (Eds.), *Taalbeheersing als tekstwetenschap: terreinen en trends*. (pp. 177-211). Dordrecht: ICG Publications.

Beaugrande, R. de (1980). *Text, Discourse and Process. Toward a multidisciplinary science of texts*. London: Longman.

Bimmel, P. & Westhoff, G. (1993). Leesstrategieën: training in het moedertaalonderwijs, transfer naar het vreemde-talenonderwijs. *Levende Talen, 485*, 577-582.

Bochardt, I. (1983). Onderzoek naar het Schrijf- & Planningsproces in de 70-er en begin 80-er jaren, op weg naar een theorie. Amsterdam: SCO-rapport 11.

Bock, J. (1982). Towards a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1-47.

Bosch, P. (1983). *Agreement and anaphora. A study of the role of pronouns in syntax and discourse*. New York: Academic Press.

Bossers, B. (1993a). VT-lezen: leesstrategieën of taalkennis? *Levende Talen, 478*, 133-139.

Bossers, B. (1993b). Nogmaals: leesstrategieën en VT-kennis. *Levende Talen, 481*, 326-330.

Bower, G., Black, J. & Turner, T. (1979). Scripts in memory for text. *Cognitive psychology, 11*, 177-220.

Breetvelt, I. (1991). *Schrijfproces en tekstkwaliteit; een onderzoek naar het verband tussen schrijfproces en tekstkwaliteit bij leerlingen in het voortgezet onderwijs*. Amsterdam: SCO-rapport 225.

Britton, J., Burgess, T., Martin, N., McLeod, A. & Rosen, H. (1975). *The development of writing abilities*. London: School Council Publications.

Brown, P. & Levinson, S. (1978). Universals in language usage: politeness phenomena. In E. Goody (Ed.), *Questions and Politeness. Strategies in social interaction* (pp.56-289). Cambridge: Cambridge University Press.

Brown, G. & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.

Butterworth, B. (1980). *Language production: Vol. 1. Speech and talk*. London: Academic Press.

Bruijn, G. de (1995). *De discourse-organiserende funktie van de vooropgeplaatste adverbiale bijzin*. Unpublished Master's thesis, University of Tilburg.

Chafe, W. (1987). Cognitive Constraints on Information Flow. In R. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21-51). Amsterdam: John Benjamins.

Christensen, F. (1967). *Notes toward a new rhetoric*. New York: Harper & Row.

Christensen, F. (1968). The problem of defining a mature style. *English Journal, 57*, 572-579.

Clifton, C. & Ferreira, F. (1987). Discourse structure and anaphora: some experimental results. In M. Coltheart (Ed.), *Attention and Performance XII* (pp. 635-654). London: Erlbaum.

Cooper, C. (1983). Procedures for describing written texts. In P. Mosenthal, L. Tamor & S. Walmsley (Eds.), *Research on writing. Principles and methods* (pp. 287-313). New York & London: Longman.

Cooper, M. & Holzman, M. (1983). Talking about protocols. *College Composition and Communication, 34*, 284-296.

Cooper, M. & Holzman, M. (1985). Reply on Counterstatement by Flower and Hayes. *College Composition and Communication 36*, 97-100.

Cooper, C. & Matsuhashi, A. (1983). A theory of the writing process. In M. Martlew (Ed.), *The psychology of written language: a developmental approach* (pp. 3-39). London: Wiley.

Daneš, F. (1974). Functional sentence perspective and the organization of the text. In F. Daneš (Ed.), *Papers on Functional Sentence Perspective* (pp. 106-128). The Hague: Mouton.

Dijk, T. van (1986). News schemata. In C. Cooper & S. Greenbaum (Eds.), *Studying writing: linguistic approaches* (pp. 155-185). Beverly Hills: Sage.

Dijk, T. van & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Dijkstra, A. & Kempen, G. (1993). *Taalpsychologie*. Groningen: Wolters-Noordhoff.

Dirksen, A. (1990). *Monitoring processes in language production.* Dissertation, University of Utrecht.

Ellis, N. (1994). *Implicit and explicit learning of languages*. New York: Academic Press.

Emig, J. (1971). *The composing processes of twelfth graders*. NCTE Research Report, 13. Urbana, Illinois: National Council of Teachers of English.

Enkvist, N. (1978). Coherence, pseudo-coherence, and non-coherence. In J. Östman (Ed.), *Reports on Text linguistics. Cohesion and semantics* (pp. 109-128). Åbo, Finland: Åbo Akademi Foundation.

Ericsson, K. & Simon, H. (1993). *Protocol Analysis. Verbal reports as data*. Cambridge, MA: MIT Press.

Faigley, L. (1981). *Using text structure models for analyzing revision.* ERIC Document reproduction Service No. ED 200 978.

Faigley, L., Cherry, R., Jolliffe, D. & Skinner, A. (1985). *Assessing writer's knowledge and processes of composing*. Norwood, N.J.: Ablex.

Faigley, L. & Witte, S. (1984). Measuring effects of revisions on text structure. In R. Beach & L. Bridwell (Eds.), *New directions in composition research* (pp. 95-108). New York: Guilford.

Frederiksen, C. (1986). Cognitive models and discourse analysis. In C. Cooper & S. Greenbaum (Eds.), *Studying writing: linguistic approaches* (pp. 227-267). Beverly Hills: Sage.

Frederiksen, C. & Dominic, J. (1981). *Writing: the nature, development, and teaching of written communication.* Vol. 2: Writing: process, development and communication. Hillsdale, NJ: Erlbaum.

Flower, L. (1979). *Problem solving strategies for writing* (third revised edition). Orlando: Harcourt Brace Jovanovich.

Flower, L. & Hayes, J. (1980). The cognition of discovery: defining a rhetorical problem. *College Composition and Communication, 31*, 21-32.

Flower, L. & Hayes, J. (1981). Plans that guide the composing process. In C. Frederiksen & J. Dominic (Eds.), *Writing; the nature, development, and teaching of written communication.* Vol. 2: Writing: process, development and communication (pp. 39-58). Hillsdale, NJ: Erlbaum.

Fox, B. (1987). *Discourse structure and anaphora.* Cambridge: Cambridge University Press.

Garrett, M. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology and learning of motivation* (pp. 133-177). New York: Academic Press.

Garrett, M. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk* (pp. 177-193). London: Academic Press.

Geest, T. van der (1991). *Tools for teaching writing as a process. Design, development, implementation and evaluation of computer-assisted writing instruction.* Dissertation, University of Twente. WMW-publikatie 7.

Geest, T. van der, Schellens, P. & Waes, L. van (1992). Het model van Flower en Hayes: een cognitief procesmodel of een retorisch taakmodel? *Tijdschrift voor Taalbeheersing 14*, 161-176.

Gernsbacher, M. (1994). *Handbook of psycholinguistics* (Ed.), New York: Academic Press.

Givón, T. (1979). *On understanding grammar.* New York: Academic Press.

Good, D. & Butterworth, B. (1980). Hesitancy as a conversational resource: some methodological implications. In H. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 145-152). The Hague: Mouton.

Graves, D. (1975). An examination of the writing process of seven-year-old children. *Research in the Teaching of English, 9*, 227-241.

Graves, D. (1981). *A case study observing the development of primary children's composing, spelling, and motor behaviours during the writing process.* New Hampshire: University of New Hampshire.

Grimes, J. (1975). *The thread of discourse.* The Hague: Mouton.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, vol. III: speech acts* (pp. 41-58). New York, Academic Press.

Grosz, S. & Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics, 12,* 175-203.

Günther, U. (1993). *Texte planen - Texte produzieren. Kognitieve Prozesse der schriftlichen Textproduktion.* Opladen: Westdeutscher Verlag GmbH.

Halliday, M. (1994). *An introduction to functional grammar.* London: Arnold.

Halliday, M. & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Harrell, L. (1957). A comparison of the development of oral and written language in school-age children. *Monographs of the Society for Research in Child Development, 22*, 1-77.

Hayes, J. (1992). Naar een nieuw model voor tekstproduktie. Een voortgangsverslag. *Tijdschrift voor taalbeheersing, 14*, 247-256.

Hayes, J. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum.

Hillocks, G. (1986). *Research on written composition. New directions for teaching*. Urbana, Illinois: ERIC/National Council of Teachers of English.

Hobbs, J. (1983). Why is discourse coherent? In F. Neubauer (Ed.), *Coherence in natural language texts* (pp. 29-70). Hamburg: Buske.

Hobbs, J. (1990). *Literature and cognition*. Menlo Park, CA: Centre for the study of Language and Information, Stanford University.

Hovy, E. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence, 63*, 341-385.

Hudson, J. & Shapiro, L. (1991). From knowing to telling: the development of children's scripts, stories and personal narratives. In A. McCabe & C. Peterson (Eds.), *Developing Narrative Structure* (pp. 89-136). Hillsdale, NJ: Erlbaum.

Hull, G. (1984). *The editing process in writing: A performance study of experts and novices*. (ERIC Document Reproduction Service no. ED 245 245).

Hunt, K. (1965). *Grammatical structures written at three grade levels*. NCTE, Research report, 3. Urbana, Illinois: National Council of Teachers of English.

Hunt, K. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development, 35*, 1-67.

Hunt, K. (1983). Sentence combining and the teaching of writing. In M. Martlew (Ed.), *The psychology of written language* (pp. 99-125). New York: Wiley.

Just, M. & Carpenter, P. (1987). *The psychology of reading and language comprehension*. Boston: Allyn and Bacon.

Kempen, G. & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*, 201-258.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construct-integration model. *Psychological Review, 95*, 163-182.

Kintsch, W. & Dijk, T. van (1978). Toward a model of text comprehension and production. *Psychological Review, 5*, 363-394.

Kolk, H. (1987). A theory of grammatical impairment in aphasia. In G. Kempen (Ed.), *Natural Language Generation. New results in artificial intelligence, psychology and linguistics* (pp. 377-391). Dordrecht: Nijhoff.

Köpper, C. & Bosshardt, H. (1994). *The influence of short-term memory and need of achievement on speech production of elderly*. Paper presented at ISAPL in Bologna (1994).

Levelt, W. (1982). Linearization in describing spatial networks. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs and questions*. Dordrecht: Reidel.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W. (1989). *Speaking; from Intention to Articulation*. Cambridge, MA: MIT Press.

Labov, W. & Waletsky, J. (1967). Narrative analysis: oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts* (pp. 12-44). Seattle: University of Washington Press.

Linde, C. & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language, 51*, 924-939.

Loban, W. (1976). *Language development: kindergarten through grade twelve*. NCTE, Research report, 18. Urbana, Illinois: National Council of Teachers of English.

Lowijck, J. & Vanmaele, L. (1992). Schrijven door jonge leerlingen: een verkenning. *Tijdschrift voor taalbeheersing, 14*, 206-220.

Lutte, G., Mönks, F., Kempen, G. & Sarti, S. (1969). *Ideaalbeelden van de Europese jeugd*. Den Bosch: Malmberg.

MacDonald, N. (1983). The UNIX Writer's Workbench Software: Rationale and Design. *The Bell System Technical Journal, 62,* 1891-1908.

Mann, W. & Thompson, S. (1987). Rhetorical Structure Theory: description and construction of text structures. In G. Kempen (Ed.), *Natural Language generation. New results in artificial intelligence, psychology and linguistics* (pp. 85-95). Boston, MA: Kluwer Academic Publishers.

Mann, W. & Thompson, S. (1988). Rhetorical Structure Theory: toward a functional theory of text organization, *Text, 8*, 243-281.

Martin, J. (1992). *English text, system and structure*. Amsterdam: Benjamins.

Matsuhashi, A. (1981). Pausing and planning: the tempo of written discourse production. *Research in the teaching of English, 15*, 113-134.

Matsuhashi, A. & Quinn, K. (1984). Cognitive questions from discourse analysis; a review and a study. *Written communication, 1*, 307-339.

McCutchen, D. & Perfetti, C. (1982). Coherence and connectedness in the development of discourse production. *Text, 2*, 113-139.

McKeown, K. (1985). *Text generation: using discourse strategies and focus constraints to generate natural language text*. Cambridge: Cambridge University Press.

Meyer, B. (1975). *The organization of prose and its effect on memory*. Amsterdam: North-Holland.

Meyer, B. (1985). Prose Analysis: purposes, procedures and problems. In B. Britton & J. Black (Eds.), *Understanding expository text*, (pp. 11-62). Hillsdale, NJ: Erlbaum.

Meyer, A. (1988). *Phonological encoding in language production: a priming study.* Dissertation, University of Nijmegen.

Mosenthal, P. (1985). Defining the expository discourse continuum. Towards a taxonomy of expository text types. *Poetics, 14*, 387-414.

Murray, D. (1978). Internal revision: a process of discovery. In C. Cooper & L. Odell (Eds.), *Research on composing: Points of departures* (pp. 85-103). Urbana, Ill: National Council of Teachers of English.

Noordman, L. & Maes, A. (1993). Tekststructuur en tekstbegrip. In A. Braet & J. van de Gein (Eds.), *Taalbeheersing als tekstwetenschap: terreinen en trends* (pp. 177-211). Dordrecht: ICG Publications.

Onrust, M., Verhagen, A. & Doeve, R. (1993). *Formuleren.* Houten: Bohn Stafleu Van Loghum.

Oversteegen, L. (1990). *Tracking time. A proposal for the representation of temporal expressions in Dutch.* Dissertation, University of Utrecht.

Paris, C. (1993). *User modelling in text generation.* London: Pinter.

Paris, C. & McKeown, K. (1987). Discourse strategies for describing complex physical objects. In G. Kempen (Ed.), *Natural Language Generation. New results in artificial intelligence, psychology and linguistics* (pp. 97-115). Dordrecht: Nijhoff.

Pianko, S. (1979). A description of the composing processes of college freshman writers. *Research on the teaching of English, 13*, 5-22.

Pool, E. van der (1993a). *A procedure for incremental structural analysis: an application to descriptive texts.* Paper presented at Xth world congress of Applied Linguistics, Aila Amsterdam: 8-10 augustus 1993.

Pool, E. van der (1993b). Schrijven met Spirit. Een bespreking van Thea van der Geest. Tools for teaching writing as a process. Design, development, implementation and evaluation of computer-assisted writing instruction. Dissertatie University of Twente 1991. *Spiegel, 11*, 95-103.

Pool, E. van der (1995a). PISA als indicator voor schrijfvaardigheid. Een toepassing op descriptieve teksten van (on)ervaren schrijvers. In A. Maes, P. van Hauwermeiren & L. van Waes (Eds), *Perspectieven in taalbeheersingsonderzoek* (pp. 204-215). Dordrecht: ICG Publications.

Pool, E. van der (1995b). Developmental changes in organizing information: a text-analytical approach. *Organization in Discourse* (in press). Åbo, Finland: Åbo Akademi Foundation.

Pool, E. van der (1995c). Developmental changes in text production: a text analytical approach. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, models and methodology. Current theories in writing research* (in press).

Pool, E. van der & Wijk, C. van (1991). Informatie in de tekst: een classificatie op basis van beschikbaarheid. In M. Bax & W. Vuijk (Eds.), *Thema's in de taalbeheersing* (pp. 151-157). Dordrecht: ICG Publications.

Pool, E. van der & Wijk, C. van (1995). Proces en strategie in een psycholinguïstisch model van schrijven en lezen. *Tijdschrift voor Onderwijs Research* (in press).

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics, 12*, 601-638.

Polanyi, L. & Scha, R. (1983). The syntax of discourse, *Text, 3*, 261-270.

Polanyi, L. & Scha, R. (1984). A syntactic approach to discourse semantics. *Proceedings of COLING 1984, 10th international conference on Computational Linguistics* (pp. 413-419). California: Stanford University.

Price, G. & Graves, R. (1980). Sex differences in syntax and usage in oral and written language. *Research in the Teaching of English, 14*, 147-153.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic Press.

Propp, V. (1968). *Morphology of the folktale*. Austin: University of Texas Press.

Ramsay, V. (1987). The functional distribution of preposed and postposed 'if' and 'when' in written discourse. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 383-408). Amsterdam: Benjamins.

Rayner, K. & Pollatsek, K. (1989). *The psychology of reading*. Englewood Cliffs: Prentice-Hall.

Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics, 29*, 1139-1172.

Reesink, G., Holleman-v.d. Sleen, S., Stevens, K. & Kohnstamm, G. (1971). Syntaktiese ontwikkeling bij schoolkinderen en volwassenen: een replicatie-onderzoek. *Nederlands Tijdschrift voor Psychologie*, 335-364.

Rentel, V. & King, M. (1983). Present at the beginning. In P. Mosenthal, L. Tamor & S. Walmsley (Eds.), *Research on writing: principles and methods* (pp. 139-176). New York: Longman.

Rumelhart, D. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: studies in cognitive science* (pp. 211-236). New York: Academic Press.

Runkel, P. & McGrath, E. (1972). *Research on human behaviour: a systematic guide to method*. New York: Holt, Reinhart & Winston.

Sanders, T. (1992). *Discourse structure and coherence*. Dissertation, University of Tilburg.

Sanders, T., Jansen, D., Pool, E. van der, Schilperoord, J. & Wijk, C. van (1995). Hierarchical structures in writing products and writing processes. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn (Eds.), *Theories, models and methodology. Current theories in writing research* (in press).

Sanders, T., Spooren, W. & Noordman, L. (1992). Towards a taxonomy of coherence relations. *Discourse Processes, 15*, 1-35.

Sanders, T., Spooren, W. & Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics, 4*, 93-133.

Sanders, T. & Wijk, C. van (1991). Tekststructuur en tekskwaliteit: van analyse naar evaluatie en revisie. In M. Bax & W. Vuijk (Eds.), *Thema's in de taalbeheersing* (pp. 140-150). Dordrecht: ICG Publications.

Sanders T. & Wijk, C. van (1995a). PISA - A procedure for analyzing the structure of explanatory texts. *Text* (in press).

Sanders, T. & Wijk, C. van (1995b). Text analysis as research tool: How hierarchical structure explains for Sentence Combining performances. In M. Levy & S. Ransdell (Eds.), *The Science of Writing.* Hillsdale, NJ: Erlbaum (to appear).

Schank, R. & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Erlbaum.

Schilperoord, J. (1995). It is about time (working title). Dissertation, University of Utrecht.

Schilperoord, J. & Pool, E. van der (1995). Tekststructuur en pauzes. Naar een cognitief model van real-time tekstproduktie. *Gramma/TTT* (to appear).

Schuurs, U. (1990). *Leren schrijven voor lezers: het effect van drie vormen van probleemgericht schrijfonderwijs op de zinsbouwvaardigheid.* Dissertation, University of Twente (WMW-publikatie-3).

Scinto, L. (1986). *Written language and psychological development.* London: Academic Press.

Smagorinsky, P. (1989). The reliability and validity of protocol analysis. *Written communication, 6*, 378-388.

Sommers, N. (1980). Revision strategies of student writers and experienced writers. *College Composition and Communication, 31*, 378-388.

Stallard, C. (1974). An analysis of writing behaviour of good students. *Research in the Teaching of English, 8*, 206-218.

Stotsky, S. (1983). Types of lexical cohesion in expository writing: implications for developing the vocabulary of academic discourse. *College Composition & Communication, 34*, 430-446.

Stotsky, S. (1990). On planning and writing plans or beware of borrowed theories. *College Composition and Communication, 41*, 37-57.

Tamor, L. & Bond, J. (1983). Text analysis: inferring process from product. In P. Mosenthal, L. Tamor & S. Walmsley (Eds.), *Research on writing; principles and methods* (pp. 99-139). New York; London: Longman.

Thomassen, A. & Galen, G. van (1992). Handwriting as a motor task: Experimentation, modelling and simulation. In: J. Summers (Ed.), *Approaches to the study of motor control and learning* (pp. 113-144). Amsterdam: Elsevier.

Thorndyke, P. (1977). Cognitive structure in comprehension and memory of narrative discourse. *Cognitive psychology, 9*, 77-110.

Tomita, M. (1986). *Efficient parsing for natural language; a fast algorithm for practical systems*. Dordrecht: Kluwer.

Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.

Vanmaele, L. & Lowyck, J. (1990). Het schrijfmodel van Flower & Hayes: een kritische analyse. *Pedagogische Studiën, 67*, 205-221.

Vendler, Z. (1967). Verbs and Times. In Z. Vendler (Ed.), *Linguistics in philosophy* (pp. 97-121). Ithaca, NY: Corneel University.

Verhagen, A. (1992). Praxis of linguistics: Passives in Dutch. *Cognitive linguistics, 3*, 301-342.

Virtanen, T. (1992). Issues of text typology: Narrative a 'basic' type of text? *Text, 12*, 293-310.

Vos, J. (1985). Schrijfprocesonderzoek en schrijfmethoden: het onderzoek van Flower en Hayes en het leerboek van Flower. *Tijdschrift voor Taalbeheersing, 7*, 258-269.

Watson, S. (1983). Syntactic change: writing development and the rethorical context. In M. Martlew (Ed.), *The psychology of written language.* (pp. 127-139). New York: Wiley.

Wilensky, R. (1983). Story grammars versus story points. *The behaviourial and brain sciences, 6*, 579-623.

Witte, S. (1987). Pre-text and composing. *College Composition and Communication, 38*, 397-425.

Witte, S. & Cherry, R. (1986). Writing processes and written products in composition research. In C. Cooper & S. Greenbaum (Eds.), *Studying writing; linguistic approaches.* (pp. 112-153). Beverly Hills: Sage.

Wijk, C. van (1987). *Speaking, writing and sentence form.* Dissertation, University of Nijmegen.

Wijk, C. van (1992). Information analysis in written discourse. In J. de Jong & L. Verhoeven (Eds.), *The construct of language proficiency* (pp. 85-99). Amsterdam: Benjamins.

Wijk, C. van (1995). *Levels of competence in writing.* Paper presented at the 5th European Workshop on Natural Language Generation. Leiden, the Netherlands.

Wijk, C. van & Kempen, G. (1982a). De ontwikkeling van syntaktische formuleervaardigheid bij kinderen van 9 tot 16 jaar. *Nederlands tijdschrift voor Psychologie, 37*, 491-509.

Wijk, C. van & Kempen, G. (1982b). Syntactische formuleervaardigheid en het schrijven van opstellen. *Pedagogische studiën*, 126-136.

Wijk, C. van & P. Lee (1995). Age-associated changes in the recall of narrative texts. Manuscript, University of Tilburg.

Wijk, C. van & Luiten, A. (1987). Zinslengte, taalvaardigheid en onderwijs. In J. Boonman & J. van der Linden (Eds.), *Taal, leren en samenwerken in het onderwijs* (pp. 5-16). Lisse: Swets & Zeitlinger.

Wijk, C. van & Pool, E. van der (1990). Standaardcorrespondentie op maat: de 'bouwsteenbrief'. *Communicatief, 3*, 150-160.

Wijk, C. van & Sanders, T. (1987). Linguïstische criteria voor tekstkwaliteit. *Levende Talen, 426*, 691-697.

Samenvatting

Schrijven als conceptueel proces

Hoe zal ik mijn tekst beginnen? Welke informatie neem ik wel op, welke toch maar niet? In welke volgorde kan ik het beste mijn mededelingen doen? Allemaal vragen die een schrijver bezig houden bij het maken van een tekst.

Naast het kiezen van de juiste woorden en het formuleren van zinnen houdt een schrijver zich bezig met het bedenken van en denken over de informatie die hij (al dan niet) in zijn tekst opneemt. Dit conceptuele deel van het schrijfproces bestaat grofweg uit *Reflecteren* op en *Selecteren* van informatie.

Bij het *Reflecteren* overweegt een schrijver zijn doelen (wat wil ik bereiken met deze tekst), en bepaalt hij een strategie (hoe kan ik dat zo goed mogelijk realiseren), waarbij ook de verwachtingen over de lezer spelen een rol spelen. Een gesteld doel bepaalt het type informatie in een tekst; bijvoorbeeld descriptieve informatie bij een informerend doel, narratieve om de lezer te amuseren, en argumentatieve informatie om te overtuigen.

Bij het *Selecteren* bepaalt de schrijver welke informatie hij al dan niet geschikt acht om op te nemen, en bedenkt een volgorde voor de op te nemen informatie. Dit Selecteer-proces wordt beïnvloed door het gestelde doel. Afhankelijk van zijn intentie, bijvoorbeeld amuseren, zal de schrijver bepaalde informatie -grappige- wel en andere -serieuze- niet opnemen. Ook de plaats van de informatie in de tekst bepaalt de schrijver aan de hand van zijn doel; wil hij een lezer bijvoorbeeld overtuigen, dan zal hij beginnen met het sterkste argument.

De precieze beslissingen die een schrijver neemt bij het Reflecteren en Selecteren zijn minder bekend. Waarin verschillen schrijvers van elkaar? Wat doet, bijvoorbeeld, een volwassen ervaren schrijver anders dan een 10-jarige beginner? Waaruit blijkt dat in de tekst? Relevante vragen die een antwoord behoeven, niet alleen om meer zicht te krijgen op menselijk taalgedrag, maar ook om te weten wat die beginner moet worden geleerd om hem vorderingen te laten maken.

Deze studie is opgezet om antwoord te krijgen op twee onderzoeksvragen: 'hoe kunnen we conceptuele processen bij het schrijven expliciteren, in regels beschrijven?' en 'hoe ontwikkelen deze processen zich met leeftijd, van 10-jarige beginnende schrijvers tot volwassen schrijvers?'.

Het schrijfproces onderzocht

Om de conceptuele processen te kunnen bestuderen moet een methodologisch probleem worden opgelost. De denkprocessen spelen zich af in de 'black box' van het menselijke brein. Het is niet mogelijk om direct toegang tot deze processen te krijgen, dus moet men een indirecte ingang vinden. Daarvoor zijn verschillende methoden. Zo kan een onderzoeker een proefpersoon tijdens het schrijven hardop laten denken, of hem na het schrijven interviewen. Deze on-line methoden vereisen echter meta-linguïstische kennis en vaardigheden van de proefpersoon. Aangezien beginnende schrijvers dergelijke kennis en vaardigheden missen, zijn deze methoden voor hen minder geschikt. Omdat we in dit onderzoek bovendien de natuurlijke wijze van tekstproduktie wilden onderzoeken, met zo min mogelijk tussenkomst van de onderzoeker, is gebruik gemaakt van een off-line methode, namelijk *tekstanalyse*.

Het is aannemelijk te veronderstellen dat teksten - de produkten van de Reflecteer- en Selecteer-activiteiten - sporen bevatten van de conceptuele processen. Analyse van deze produkten biedt een basis om (aspecten van) conceptuele processen te herleiden. De vraag is dan natuurlijk wèlke sporen een tekst bevat, en hoe ze inzicht kunnen bieden in de activiteiten die hebben plaatsgehad. We zullen dit illustreren aan de hand van twee voorbeeldteksten, gepresenteerd in (1) en (2). Ze zijn afkomstig uit het corpus dat in het onderzoek gebruikt is. Het corpus bestond uit 86 teksten geschreven door kinderen van 10-, 12- en 15 jaar en volwassenen.

De tekst in (1) is geschreven door een 10-jarige jongen (en bevat spelfouten); de tekst in (2) is geschreven door een volwassene met een universitaire achtergrond. De opdracht was een persoon te beschrijven op wie de schrijver graag wilde lijken, en iets te vertellen over zijn of haar personalia, karakter, uiterlijk, beroep en hobbies. De tekst is gesegmenteerd in deelzinnen die genummerd zijn.

(1) (1) Hij is onderwijzer in Heervarebeek. (2) Hij leert kinderen leren (2a) zoals reken Taal Geschiedenis Aardrijkunde. (3) Hij is zoiets 40 jaar. (4) Hij ziet er met een pak uit en een witte bloes en zwarte schoenen en gezond. (5) Kinderen leren. (6) Zondags Woensdag Nademiddig en Zaterdag Na de Middag fietsen. (6a) schriften Na kijken. (6b) Wandelen in de bossen. (7) Hij heekt Oome harie. (8) Zoms gaat hij naar tilburg kijken Naar de familis (8a) en ook winkels kijken Zondags.

(2) (1) Ik zou (1a) als het dan toch moet (1) wel op Ed Nijpels willen lijken. (2) Hij ziet er best leuk uit (3) is goed gekleed (4) en kan prima uit zijn woorden komen. (5) Bovendien heeft hij allerlei activiteiten ondernomen naast zijn, toch ook redelijk afwisselende, loopbaan. (6) Juist deze afwisseling trekt me erg aan. (7) Zijn

"hoofdfunctie" op dit moment is het burgemeesterschap van Breda, (8) maar als presentator op de televisie staat hij ook zijn mannetje! (9a) Als hij zich ontspant na zijn drukke bezigheden, (9) is hij op de tennisbaan te vinden (10) en 's avonds is hij graag samen met vrienden of juist lekker alleen met een fijn boek! (11) Door zijn open karakter maakt hij makkelijk vrienden (12) en legt hij makkelijk contacten. (13) Als burgemeester probeert hij dicht bij de bevolking te staan. (14) De leeftijd van Ed Nijpels (14a) (zo rond de 40) (14) is een leeftijd die veel voordelen heeft: (15) je hebt de kans gehad (en de tijd!) om iets van je leven te maken, (16) je hoeft niet meer zo nodig de 'stoere bink' uit te hangen. (17) Samenvattend moet ik zeggen dat het in ieder geval lijkt alsof Ed Nijpels zichzelf prettig voelt, (18) of (18a) om het populair te zeggen (18) 'goed in zijn vel zit'. (19) En dàt is altijd iets om na te streven!

Als we de twee teksten bekijken, zien we een aantal verschillen in onder andere woordkeus, grammaticaliteit, hoeveelheid informatie, en gebruik van interpunctie. Uit de hoeveelheid gepresenteerde informatie - te meten als tekstlengte - zou je bijvoorbeeld kunnen afleiden dat de schrijver van de eerste tekst minder kennis heeft dan die van de tweede tekst. Een dergelijke intuïtieve interpretatie is natuurlijk niet objectief. Bovendien is deze gebaseerd op eerste indrukken en niet op een geëxpliciteerde analyse.

Voor een betrouwbaar resultaat moet duidelijk zijn welke tekstkenmerken worden geanalyseerd; bovendien moet de analyse volgens expliciet geformuleerde regels verlopen.

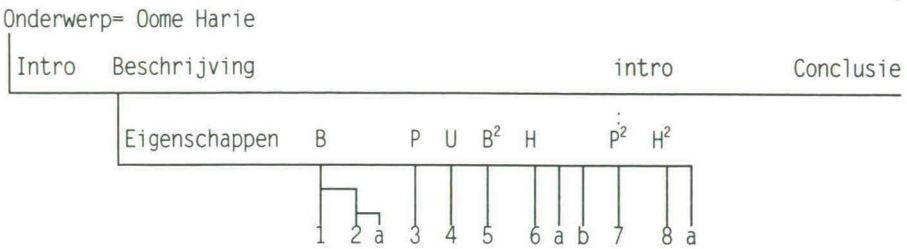
Uit eerder onderzoek is gebleken dat de structuur en inhoud van een tekst de conceptuele processen het beste weerspiegelen. De tekststructuur, in combinatie met de inhoud, is het eindresultaat van het - al dan niet bewust - ordenen van informatie. Dat is de belangrijkste reden om in dit onderzoek te kiezen voor een analyse van de *tekststructuur*.

PISA: een methode voor structuur-analyse

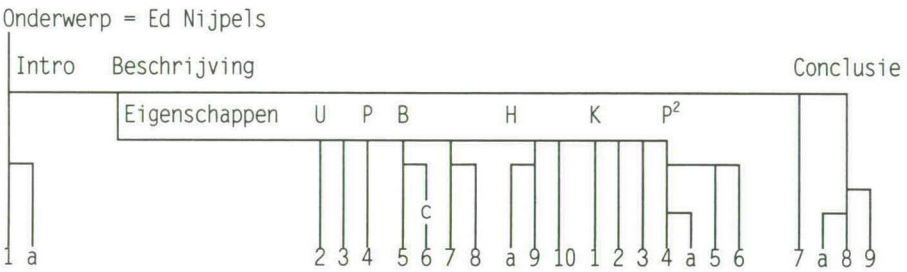
Aangezien bestaande analysemethoden onvoldoende expliciet geformuleerd waren, is een nieuwe procedure ontwikkeld om de structuur van een tekst te analyseren. De methode, genaamd PISA, een acroniem voor Procedures voor Incrementele Structuur Analyse, is gebaseerd op methoden en theoretische inzichten uit, onder andere, de (psycho)linguïstiek. In de analyse wordt iedere (deel)zin geïnspecteerd op de aanwezigheid van indicatoren, zoals voegwoorden en verwijswoorden, die de samenhang in een tekst bepalen. Op basis van deze tekstkenmerken worden de (deel)zinnen aan elkaar verbonden en wordt de structuur zichtbaar.

Als we PISA toepassen op de voorbeeldteksten, dan levert dat de onderling vergelijkbare structuren op zoals gepresenteerd in figuur 1 en 2. De namen van de thema's in de persoonsbeschrijving - overeenkomstig de opdracht - zijn afgekort op de Eigenschappen-lijn.

De afkortingen staan voor: Personalia (afgekort tot *P*), Karakter (*K*), Uiterlijk (*U*), Beroep (*B*), Hobby (*H*). Het superschrift '^{2}' betekent hervatting van een thema; de 'c' in figuur 2 duidt op een expliciet gegeven commentaar van de schrijver op de beschreven persoon.



Figuur 1. Structuur van de eerste voorbeeldtekst gepresenteerd in (1)



Figuur 2. Structuur van de tweede voorbeeldtekst gepresenteerd in (2)

Een PISA-analyse geeft op verschillende niveaus de structuur van een tekst weer. De structuren informeren ons over:

- de globale structuur, *superstructuur* genaamd. Deze geeft aan dat de beschrijving van de tweede tekst wordt voorafgegaan door een inleiding en afgesloten met een conclusie; bij de eerste tekst is dat niet het geval.
- de thematische indeling van de eigenschappen die worden beschreven, *macrostructuur* genaamd. In de eerste tekst wordt niets over het karakter verteld in tegenstelling tot de tweede tekst die alle vijf de thema's beschrijft; de eerste tekst bevat drie keer een hervatting van een thema, de tweede tekst is slechts één keer thematisch discontinu.
- de hiërarchische positie van opeenvolgende deelzinnen, *microstructuur* genaamd. In de

eerste tekst zijn de meeste deelzinnen hiërarchisch nevengeschikt, behalve segment 2a dat ondergeschikt is aan het voorgaande; in de tweede tekst zijn 1a, 6, 9a, 14a-16, 18a en 19 ondergeschikt, de andere zijn nevengeschikt.

Daarnaast is de betekenis van iedere verbinding gelabeld; zo is bijvoorbeeld deelzin 1a in de tweede tekst een voorwaarde bij 1.

TRACE: een reconstructie van Reflecteer- en Selecteer-processen

De bovenstaande PISA-structuren beschrijven de teksten, echter nog niet de conceptuele processen die daaraan ten grondslag liggen. We kunnen de Reflecteer- en Selecteer-processen er wel gedeeltelijk uit afleiden.

De PISA-structuren zijn op te vatten als resultaten van, onder andere, Reflecteer- en Selecteer-processen en zijn daarom bruikbaar om delen van die processen te *reconstrueren*. De beslissingen die een schrijver heeft genomen (welk thema bespreekt hij, hoe werkt hij dit uit, leidt hij zijn onderwerp in en uit, etc.) zijn te herleiden uit de structuren en vervolgens te beschrijven in de vorm van herschrijfregels. Zo kunnen processen die hebben geleid tot de deelzinnen 1 t/m 3 in de eerste tekst worden beschreven als in tabel 1. De reconstructieregels staan links, de deelzinnen rechts.

Tabel 1. Reconstructie van Reflecteer en Selecteer-activiteiten bij het schrijven van de deelzinnen 1 t/m 3 van de eerste tekst

Reconstructieregel		Deelzin
KIES EEN THEMA →	Beroep	1
WERK THEMA UIT →	Specificeer eigenschap	2
WERK THEMA UIT →	Specificeer eigenschap	2a
KIES EEN THEMA →	Personalia	3

Voor alle teksten van het corpus zijn op basis van de PISA-structuren (aspecten van) de conceptuele processen gereconstrueerd en in regels beschreven zoals in tabel 1. Dit heeft geresulteerd in een TRACE-model voor het schrijven van een persoonsbeschrijving. TRACE is een acroniem voor Tekst-gebaseerde Reconstructies van Activiteiten in het ConceptuelE deel van het schrijfproces. Het TRACE-model bestaat uit een set van regels die corresponderen met activiteiten die uitgevoerd worden tijdens het schrijven van een tekst. De set bevat twee hoofdregels: een BESCHRIJF-regel, corresponderend met Selecteer-activiteiten, en een

EVALUEER-regel, corresponderend met Reflecteer-activiteiten. Deze hoofdregels roepen ieder weer ingebedde regels aan. Als we de TRACE-regels toepassen op de twee voorbeeldteksten krijgen we de reconstructies zoals gepresenteerd in tabel 2 en 3.

Tabel 2. Reconstructie van de conceptuele activiteiten bij het schrijven van een persoonsbeschrijving gebaseerd op de PISA-structuur in figuur 1. De reconstructie levert als output de deelzinnen van de tekst op.

TRACE-regel		Deelzin
BESCHRIJF Onderwerp		
→ **INTRODUCEER Onderwerp**		
+ **KARAKTERISEER Onderwerp**		
KARAKTERISEER Onderwerp*		
→ KIES THEMA	→ Beroep	1
→ WERK THEMA UIT	→ Specificeer eigenschap	2
→ WERK THEMA UIT	→ Specificeer eigenschap	2a
→ KIES THEMA	→ Personalia	3
→ KIES THEMA	→ Uiterlijk	4
→ KIES THEMA	→ Beroep	5
→ KIES THEMA	→ Hobby	6
→ KIES THEMA	→ Hobby	6a
→ KIES THEMA	→ Hobby	6b
INTRODUCEER Onderwerp	→ Noem naam	7
KARAKTERISEER Onderwerp		
→ KIES THEMA	→ Hobby	8
→ KIES THEMA	→ Hobby	8a

'*' betekent recursief (herhaald) gebruik

De reconstructie in tabel 2 laat zien dat de schrijver een beperkte set van activiteiten heeft uitgevoerd voor het schrijven van zijn tekst. Steeds gebruikt hij dezelfde procedure (KIES THEMA) om zijn onderwerp te karakteriseren; slechts twee keer werkt hij een genoemde eigenschap uit (WERK THEMA UIT die resulteren in segment 2 en 2a). De volgorde waarin de schrijver van deze tekst de regels toepast, is opvallend. Hij begint niet met het introduceren van het onderwerp, maar met het noemen van eigenschappen. Nadat hij enkele eigenschappen heeft beschreven, noemt hij pas de naam van de persoon in kwestie, waarna hij vervolgt met een karakterisering van Oome Harie.

Tabel 3. Reconstructie van conceptuele activiteiten bij het schrijven van de tekst in (2) op basis van de PISA-structuur in figuur 2. De reconstructie levert als output de deelzinnen van de tekst op.

TRACE-regel		Segment
EVALUEER (BESCHRIJF Onderwerp)		
→ **EVALUEER (INTRODUCEER Onderwerp)**		
+ **EVALUEER (KARAKTERISEER Onderwerp)**		
+ **EVALUEER Beschrijving**		
EVALUEER (INTRODUCEER Onderwerp)		
→ INTRODUCEER Onderwerp	→ Noem naam	1
→ EVALUEER (INTRODUCEER)	→ Negatieve attitude	1a
EVALUEER (KARAKTERISEER Onderwerp)		
→ KARAKTERISEER Onderwerp		
→ KIES THEMA	→ Uiterlijk	2
→ KIES THEMA	→ Uiterlijk	3
→ KIES THEMA	→ Personalia	4
→ KIES THEMA	→ Beroep	5
→ SPECIFICEER-EVALUATIE	→ Positieve opinie	6
→ KARAKTERISEER Onderwerp		
→ KIES THEMA	→ Beroep	7
→ WERK THEMA UIT	→ Specificeer eigenschap	8
→ GEEF CONTEXT	→ Conditie	9a
→ KIES THEMA	→ Hobby	9
→ KIES THEMA	→ Hobby	10
→ KIES THEMA	→ Karakter	11
→ KIES THEMA	→ Karakter	12
→ KIES THEMA	→ Karakter	13
→ KIES THEMA	→ Personalia	14
→ WERK THEMA UIT	→ Specificeer eigenschap	14a
→ WERK THEMA UIT	→ Specificeer eigenschap	15
→ WERK THEMA UIT	→ Specificeer eigenschap	16
EVALUEER (Onderwerp en Beschrijving)		
→ **EVALUEER (Onderwerp)**		
→ Vat samen		17
→ Vat samen		18a
→ Vat samen		18
→ **EVALUEER (Beschrijving)**		19

De reconstructie in tabel 3 laat zien dat deze schrijver meer verschillende regels heeft gebruikt. Naast de BESCHRIJF-regels heeft hij ook EVALUEER-regels gebruikt. Zo geeft hij zijn negatieve attitude ten aanzien van de opdracht (resulterend in segment 1a), en geeft zijn positieve opinie

weer ten aanzien van de in het vorig segment beschreven eigenschap (segment 6). Bij het toepassen van de BESCHRIJF-regels, werkt de schrijver een aantal keren een thema uit (segment 8, 14a-16); in één geval plaatst hij een eigenschap in een context (segment 9a). In tegenstelling tot de eerste schrijver, besluit deze schrijver zijn tekst met een korte samenvatting en maakt hij een afsluitende opmerking over de opdracht (segment 19).

Verschillen tussen beginnende en ervaren schrijvers

Deze reconstructies hebben we gebruikt om inzicht te krijgen in de precieze beslissingen die een schrijver neemt, maar ook om de ontwikkeling van de conceptuele processen met leeftijd in kaart te brengen. De TRACE-regels onthullen (voor een deel) hoe de schrijvers te werk zijn gegaan. Zo laten de voorbeelden zien dat de eerste schrijver alleen BESCHRIJF-regels heeft gebruikt, terwijl de tweede schrijver ook EVALUEER-regels heeft gebruikt.

Voor alle schrijver van de corpusteksten is een TRACE-tabel gemaakt. Dit leverde 86 verschillende karakteriseringen op van de individuele schrijvers, die echter bij vergelijking regelmatige patronen vertoonden in het gebruik van de regels. Het bleek dat verschillende schrijvers dezelfde regels toepasten (weliswaar met volgordevarianten).

Op basis van deze overeenkomsten in het gebruik van de regels hebben we twee onderliggende strategieën afgeleid die ten grondslag liggen aan het geven van een persoonsbeschrijving: een Beschrijf-strategie, waarin alleen de BESCHRIJF-regels worden toegepast, en een Evalueer-strategie, waarbij de BESCHRIJF-regels worden aangevuld met EVALUEER-regels. De schrijver van de eerste tekst, is een voorbeeld van iemand die de Beschrijf-strategie hanteert, de schrijver van de tweede tekst, van iemand die de Evalueer-strategie hanteert.

Om de ontwikkeling met leeftijd na te gaan is bekeken in hoeverre de strategieën correspondeerden met leeftijd. Het bleek dat 10- en 12-jarigen vrijwel alleen de Beschrijf-strategie hanteerden; het merendeel van de 15-jarigen en alle volwassenen pasten de Evalueer-strategie toe. Deze verschuiving van de Beschrijf- naar de Evalueer-strategie betekent een uitbreiding van het gebruikte regel-repertoire. Naast deze kwalitatieve verandering met leeftijd was er ook een kwantitatief verschil te constateren. Oudere schrijvers gebruikten per tekst meer BESCHRIJF-regels dan de jongere. Daarnaast lieten de volwassenen, in tegenstelling tot de andere schrijvers, een geordend gebruik van de regels zien; de besprekingen van de thema's waren niet of nauwelijks discontinu.

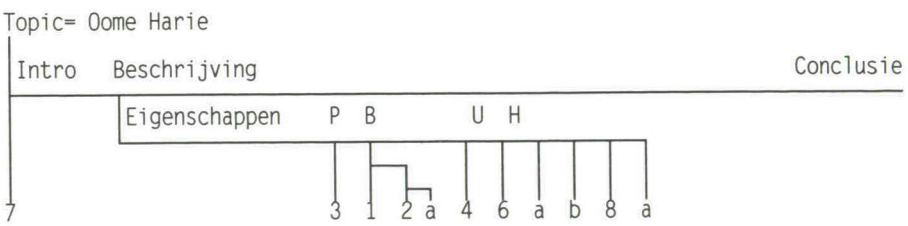
Planning en revisie in onderwijs en training

De resultaten van deze studie kunnen van nut zijn bij schrijfonderwijs aan kinderen en schrijftrainingen voor volwassenen. Belangrijk is dat bij het vergroten van de

schrijfvaardigheid niet alleen gelet wordt op formuleringskwesties, maar op de dieperliggende conceptuele aspecten, het Reflecteren op en het Selecteren van informatie. Het kiezen van de inhoud en het structureren daarvan bepalen voor een groot gedeelte de leesbaarheid - en daarmee de kwaliteit - van een tekst. Onderwijs en training zouden erop gericht moeten zijn een schrijver zich hiervan bewust te maken door de aandacht voor het werkelijke schrijven uit te breiden met aandacht voor planning en revisie. Een schrijver kan problemen voorkomen of oplossen door vooraf na te denken over welke informatie in de tekst moet komen en die schematisch te ordenen, dus door te *plannen*, en door na het schrijven zijn tekst te *reviseren* door de inhoud en structuur kritisch te bekijken. Bij een dergelijke aanpak staat het conceptuele proces centraal.

Bij planning spelen het doel en de strategie-keuze een bepalende rol. De ontwikkeling van een schrijver kenmerkt zich het meest door uitbreiding van het repertoire aan strategieën. Dit impliceert echter dat een schrijver, om vorderingen te kunnen maken, zich verschillende strategieën zich moet kunnen maken om ze toe te passen.

De resultaten van dit onderzoek kunnen verder een aanzet vormen tot het reviseren van teksten. Op basis van de PISA-structuren is het mogelijk met marginale organisatiewijzigingen de tekst te verbeteren. Zo zagen we in de voorbeeldtekst over Oome Harie in (1) dat de schrijver zijn onderwerp pas in deelzin 7 introduceerde hetgeen de beschrijving onderbrak. De schrijver hernam bovendien thema's enkele keren, hetgeen een incoherente presentatie tot gevolg had. Aangezien de volgorde van de informatie cruciaal is voor de begrijpelijkheid van de tekst, is het een criterium voor tekstkwaliteit. Voor de voorbeeldtekst (1) kan een eenvoudige revisie de tekst enigszins verbeteren. Door redundante deelzinnen te verwijderen (deelzin 5 die een herhaling is van 2) en deelzinnen thematisch te ordenen (deelzin 7 naar voren plaatsen, gevolgd door deelzin 3). Deze verschuivingen leveren een thematisch gestructureerde tekst op. De structuur en de tekst staan in figuur 3 en in (3).



Figuur 3. Structuur van de herziene voorbeeldtekst zoals gepresenteerd in (3)

(3) (7) Hij heekt Oome harie. (3) Hij is zoiets 40 jaar. (1) Hij is onderwijzer in Heervarebeek. (2) Hij leert kinderen leren (2a) zoals reken Taal Geschiedenis Aardrijkunde. (4) Hij ziet er met een pak uit en een witte bloes en zwarte schoenen en gezond. (6) Zondags Woensdag Nademiddig en Zaterdag Na de Middag fietsen. (6a) schriften Na kijken. (6b) Wandelen in de bossen. (8) Zoms gaat hij naar tilburg kijken Naar de familis (8a) en ook winkels kijken Zondags.

Deze gereviseerde tekst is natuurlijk nog steeds verre van perfect. De tekst wordt niet expliciet afgesloten met een conclusie, en de persoonsbeschrijving dekt niet alle vijf de thema's. Maar het voorbeeld laat zien dat met marginale structuurwijzigingen en inachtneming van het niveau van de schrijver het mogelijk is een tekst te reviseren. Het geeft ook aan dat een schrijver zich de vragen 'hoe zal ik mijn tekst beginnen?', 'welke informatie neem ik wel en niet op?' en 'in welke volgorde kan ik het beste mijn mededelingen doen?' ook ná het schrijven van de tekst kan stellen voor een beter resultaat.

Appendix to chapter 1: the original Dutch texts

The Dutch versions of the texts presented in (1) and (2) in chapter 1.

(1) Hij is onderwijzer in Heervarebeek. Hij leert kinderen leren zoals reken Taal Geschiedenis Aardrijkunde. Hij is zoiets 40 jaar. Hij ziet er met een pak uit en een witte bloes en zwarte schoenen en gezond. Kinderen leren. Zondags Woensdag Nademiddig en Zaterdag Na de Middag fietsen schriften Na kijken Wandelen in de bossen. Hij heekt Oome harie. Zoms gaat hij naar tilburg kijken Naar de familis en ook winkels kijken Zondags.

(2) Ik zou als het dan toch moet wel op Ed Nijpels willen lijken. Hij ziet er best leuk uit is goed gekleed en kan prima uit zijn woorden komen. Bovendien heeft hij allerlei activiteiten ondernomen naast zijn, toch ook redelijk afwisselende, loopbaan. Juist deze afwisseling trekt me erg aan. Zijn "hoofdfunctie" op dit moment is het burgemeesterschap van Breda, maar als presentator op de televisie staat hij ook zijn mannetje! Als hij zich ontspant na zijn drukke bezigheden, is hij op de tennisbaan te vinden en 's avonds is hij graag samen met vrienden of juist lekker alleen met een fijn boek! Door zijn open karakter maakt hij makkelijk vrienden en legt hij makkelijk contacten. Als burgemeester probeert hij dicht bij de bevolking te staan. De leeftijd van Ed Nijpels (zo rond de 40) is een leeftijd die veel voordelen heeft: je hebt de kans gehad (en de tijd!) om iets van je leven te maken, je hoeft niet meer zo nodig de 'stoere bink' uit te hangen. Samenvattend moet ik zeggen dat het in ieder geval lijkt alsof Ed Nijpels zichzelf prettig voelt, of om het populair te zeggen 'goed in zijn vel zit'. En dàt is altijd iets om na te streven!

Appendix to chapter 4: the procedures and functions of PISA

Domain of application

PISA has been developed for expository texts. This genre consists of several subtypes, among them explanations and descriptions. The first subtype is central in research by Sanders and Van Wijk (1995a,b), the second one in the present study. The main difference between these two subtypes is that descriptions are dominated by a listing of static predications, whereas explanations very often also contain sequences of events or actions. PISA, as presented in this appendix, is capable of analyzing both these listings and sequences, and therefore is more powerful than strictly needed for the person descriptions analyzed in this study.

In the analysis of descriptive texts the INSPECT-functions TEMPORAL-MARKER and RESPONSE-MARKER are almost never used and ACTION-MARKER and MODAL-MARKER rarely. Of course this also goes for the CONNECT-rules depending most heavily on these functions (Response-1 to Response-3; Action-1 and Action-2).

Main characteristics

PISA is a text-analytic instrument that assigns a labelled hierarchical structure to an expository text on the basis of explicitly mentioned linguistic elements and a shallow data-base. The instrument is made up by three functions: INSPECT, CONNECT, and INTEGRATE.

PISA accepts as input discourse segments, in most cases a single main clause or subordinated clause. Each segment has been parsed, i.e., the lexical meanings and syntactic functions of its constituents are specified. In each segment the constituents are placed in the English word order, i.e., inversions of subject and verb by a fronted adverbial phrase are undone and split verbal groups are rejoined (see Sanders & Van Wijk, 1995a).

INSPECT checks each segment for the presence of a selected set of linguistic elements. Many of these elements are literally present (e.g., auxiliary verbs); they have been incorporated as a full listing. Some elements, however, ask for a more complex retrieval operation (e.g., to notice that being a butcher is someone's profession); these have been made available in a shallow database. For each element found, INSPECT may specify up to three features: *type*, *token*, and *position*. This information is transferred partly to CONNECT, partly to INTEGRATE.

CONNECT consists of an ordered set of condition-action pairs. On the basis of INSPECT-information it is determined which conditions apply, and by that which actions have to be executed. Thus, CONNECT yields for each segment a decision with respect to its position in the microstructure and the relational meaning of this connection.

INTEGRATE is made up by condition-action pairs as well. On the basis of INSPECT-information it decides on the macro- and superstructure of the text.

Procedures

Name: **INSPECT**

Aim: INSPECT examines each discourse segment for the presence of specific linguistic elements. This is done by applying nine functions. Each time a function fires, i.e. it has found an instance of the elements it has to look for, the procedure assigns a *value* to the *features* 'type', 'token', and 'position' of this element (if relevant for this specific case). In addition, INSPECT searches an antecedent for endophoric referential expressions.

'Type' indicates the linguistic category of the element (e.g., adverb, word-stem or negation) or the route of recovering the element (direct, derived).

'Token' represents a specific meaning of the element (e.g. *sometimes* denotes a frequency and *if* a condition).

'Position' specifies the place of the element in syntactic surface structure: *preverb* if element precedes main and/or auxiliary verb, *verb* if element happens to be a main or auxiliary verb, and *postverb* if element follows main and/or auxiliary verb.

Input: Lexically and syntactically parsed discourse segments.

Output: Lists of the elements present in each segment supplemented with values for the features type, token and position. The output of functions I1 to I7 is sent to CONNECT, the output of functions I8 and I9 to INTEGRATE.

Rules: See table 1 for a specification of the INSPECT-functions; they belong to three clusters that correspond to the typology in section 4.5, table 7:

a. functions that examine the segment for elements relevant for the microstructure (I1 to I7);
b. a function that examines the segment for elements relevant for the macro structure (I8);
c. a function that examines the segment for elements relevant for the superstructure (I9).

See table 2 for a definition of these functions.

Name: **CONNECT**

Aim: CONNECT assigns to each discourse segment a labelled hierarchical position in the microstructure. The procedure consists of a set of condition-action pairs. The feature-values pairs, resulting from INSPECT, are tested in the condition part of a rule. The condition-action pairs are run through in a fixed order, corresponding to their priority order in determining a text's structure. The order is as follows: presence of a contraction, a connector, a lexical marker (response, modal, action, and temporal), and finally a referential expression.
CONNECT makes use of three auxiliary variables: $Context, $Segment, and $Time. These are defined in table 3.

Input: Feature-value pairs assigned by INSPECT-functions.

Output: Actions listed in their order of acceptance, i.e., the one decided on first is placed in front ('first in - first out').
From this list, the primary candidate for a micro-structure link is choosen as follows:
- If no actions appear on the list, regard segment as unconnectable.
- If only one action appears on the list, select it.
- If there is more than one action on the list and the first action is 'collocate segment', select the action coming second in the list.
- In all other cases, select the first action.

Rules: See table 4 for a definition of the function. These rules can be grouped in four clusters that correspond to the most prominent feature in their condition part (in most cases mentioned first). These are subsequently:
a. rules which base their actions mainly on syntactic form;
b. rules which base their actions mainly on connectors;
c. rules which base their actions mainly on lexical markers;
d. rules which base their actions mainly on referential expressions.

Name: **INTEGRATE**

Aim: INTEGRATE links segments to the macro- and superstructure on the basis of the *function* they have in common, e.g. introducing a person, describing the protagonist with respect to specific themes, commenting on it, and concluding the description. It assigns a superstructure and macrostructure on the basis of the INSPECTed elements and integration with world and discourse knowledge.
In addition it checks the linkings made in to the different levels of the structure. When a segment has not been attached, has a discontinuous linking to an other segment, or an isolated linking to a line in the Macrostructure (for instance to the Action-line), INTEGRATE replaces the first choice for a linking by the next entry on the list which resulted from running through CONNECT.

Input: Feature-value pairs assigned by INSPECT-functions.

Output: A macrostructure, a superstructure, and a check of each segment's linking in the microstructure.

Rules: See table 5 for a specification of these functions.

Table 1. Definition of INSPECT-rules

Code	Function	Condition	Action
I1	CONTRACT	true	assign value to <position>: {preverb, verb, postverb}
I2	CONNECTOR	true	assign value to <position>: {main clause, initial subclause, final subclause, medial subclause} <type>: {subordinator, coordinator, adverb} <token>: {causal, conditional, temporal...}
I3	REFERENT	true	assign value to <position>: {preverb, verb, postverb} <type>: {full NP, pronoun, word-stem} <token>: {topic, deictic, endophor} <antecedent-position-in-text>: {adjacent, remote} <antecedent-position-within-segment>: {preverb, verb, postverb, segment}
I4	ACTION-MARKER	true	assign value to <type>: {auxiliary verb, main verb} <token>: {action, intension, sensation, immobility, circumstance}
I5	MODAL-MARKER	true	assign value to <type>: {auxiliary verb, adverb} <token>: {obligation, desire, possibility}
I6	TEMPORAL-MARKER	true	assign value to <type>: {lexical, verb-inflexion} <token>: {sequence, frequency, overlap, punctual, past}
I7	RESPONSE-MARKER	true	assign value to <type>: {lexical, negation} <token>: {problem, goal}
I8	THEME-MARKER	true	assign value to <type>: {direct, derived} <token>: {appearance, character, hobby, career, personal particulars}
I9	SUPER-STRUCTURE MARKER	true	assign value to <type>: {direct, derived} <token>: {introduction, description, conclusion, comment}

Table 2. Definition of INSPECT-functions

Code	Function	Definition
I1	CONTRACT	Inspects a segment for contracted constituents. If this is the case, the function assigns a value to the feature <position>. **<position=preverb>**: contracted element precedes main verb **<position=verb>**: main verb has been contracted **<position=pre+verb>**: main verb and its precessor have been contracted **<position=postverb>**: contracted element follows main verb If there is no contraction, the function returns false.
I2	CONNECTOR	Inspects a segment for the presence of a connector. If this is the case, the function assigns a value to the features <position>, <type> and <token>. **<position=main clause>**: connector in a main clause **<position=initial subclause>**: connector in an initial subclause **<position=medial subclause>**: connector in a medial subclause **<position=final subclause>**: connector in a final subclause **<type=subordinator>**: connector is a subordinating conjunctor **<type=coordinator>**: connector is a coordinating conjunctor **<type=adverb>**: connector is conjunctive adverb(ial phrase) **<token=causal>**: connector indicates a causality (*because*) **<token=conditional>**: connector indicates a condition (*if*) **<token=temporal>**: connector indicates a temporal aspect (*when*) **<token=purpose>**: connector indicates a purpose (*in order to*) **<token=additive>**: connector indicates an addition (*and*) **<token=alternative>**: connector indicates an alternative (*or*) **<token=concession>**: connector indicates another relation (*but*) If there is no connector, the function returns false.
I3	REFERENT	Inspects a segment for the presence of a (nontemporal) referential expression. If this is the case, the function assigns a value to the features <position>, <type>, and <token>. For endophoric references it also assigns a value to the features <position-antecedent-in-text> and <position-antecedent-in-segment>. **<position=preverb>**: referential expression precedes main verb **<position=verb>**: referential expression is main verb **<position=postverb>**: referential expression follows main verb **<type=full NP>**: referential expression is a full NP or a proper name. **<type=pronoun>**: referential expression is a personal pronoun or an independently used possessive or demonstrative pronoun **<type=word-stem>**: referential expression shares a lexical stem with a word in a preceding segment **<token=topic>**: reference to the discourse topic (e.g. *he* to protagonist) **<token=deictic>**: reference to a discourse participant (e.g. *me* to writer) **<token=endophor>**: reference to an object or a person mentioned earlier in the text (and not being the topic)

<antecedent-position-in-text=adjacent>: antecedent is present in the directly preceding segment; line-number of this segment is assigned to auxiliary variable $Segment

<antecedent-position-in-text=remote>: antecedent is present in a segment before the directly preceding one; line-number of this segment is assigned to the auxiliary variable $Segment

<antecedent-position-in-segment=preverb>: antecedent precedes main verb

<antecedent-position-in-segment=verb>: antecedent is main verb (and not a copula verb)

<antecedent-position-in-segment=postverb>: antecedent follows the main verb

<antecedent-position-in-segment=segment>: antecedent is predicate or even entire segment

If there is no referent, the function returns false.

I4 ACTION-MARKER

Inspects a segment for the presence of an action-marker.
If this is the case, the function assigns a value to the features <type> and <token>.

<type=main verb>: element is a main verb

<type=auxiliary verb>: element is an auxiliary verb

<token=action>: main verb or auxiliary verb denotes an action (*walk, eat; goes x-ing, starts to x*)

<token=epistemic>: main verb denotes an epistemic experience (*think, believe*)

<token=sensation>: main verb denotes a sensory experience (*see, feel*)

<token=immobility>: main verb denotes an immobility (*lie, sit*)

<token=circumstance>: main verb denotes a circumstance (*wear, possess*)

If there is no main verb, the function returns false.

I5 MODAL-MARKER

Inspects a segment for the presence of a modal-marker.
If this is the case, the function it assigns a value to the features <type> and <token>.

<type=adverb>: element is an adverb

<type=auxiliary verb>: element is an auxiliary verb

<token=obligation>: element denotes an obligation (*have to, must, need to*)

<token=desire>: element denotes a desire (*want, like to*)

<token=possibility>: element denotes a possibility (*can, may*).

If there is no modal-marker, the function returns false.

I6 TEMPORAL-MARKER

Inspects a segment for the presence of a temporal-marker.
If this is a case, the function assigns a value to <type> and <token>.

<type=lexical>: element is an adverb (*then, after that*), a prepositional phrase (*on december 5th*), or a verb with an inchoative or telic meaning (*to leave, to arrive*)

<type=verb-inflexion>: element is a verb inflexion (*has* lived)

<token=sequence>: element denotes a sequentiality (*then, afterwards, subsequently*)

<token=frequency>: element denotes a frequency (*sometimes, often, regularly, always, every year*)

<token=overlap>: element denotes an overlap (*at the same time, in the meanwhile*)
<token=punctual>: element denotes a punctual time (*on december 5th*)
<token=past>: element denotes an event in the past (e.g. *has slept*)
If there is no temporal-marker, the function returns false.

I7 RESPONSE-MARKER

Inspects a segment for the presence of a response-marker.
If this is the case, the function assigns a value to <type> and <token>.
<type=lexical>: element is a word or phrase denoting a problem (*illness, accident*) or a goal (*intend*)
<type=negation>: element is negation (*not, no*) or a negative prefix (*unhappy*)

<token=problem>: element denotes a problem (*feel uneasy, be ill*)
<token=goal>: element denotes a goal (*strive for, intent*)
If there is no response-marker, the function returns false.

I8 THEME-MARKER

Inspects a segment for the presence of theme-markers.
If this is the case, the function assigns for each marker a value to the features <type> and <token>.
<type=direct>: element refers litterally to a theme in the macrostructure
<type=derived>: element refers to a theme by a reasoning on the basis of world knowledge

<token=character>: element is a word or phrase expressing an aspect of a person's character (e.g. *character, he is* + characteristic)
<token=appearance>: element is a word or phrase expressing an aspect of the protagonist's appearance (e.g. *appearance, have/be*+a physical part of the body, *to wear*, the length, physical objects or parts in the subject position (his hair is voorovergekamd, his cloths...))
<token=career>: element is a word or phrase expressing an aspect of the protagonist's profession (e.g. *the/his profession, to work at, ...*)
<token=hobby>: element is a word or phrase expressing an aspect of the protagonist's hobby (e.g. *relaxation, hobby, time off/spare time, he likes to do, ...*)
<token=personal particulars>: element is a word or phrase expressing an aspect of the protagonist's personal particulars (e.g. *his age, he lives, to be born, to be+civil state, family/other relatives, to owe non-physical objects, smoking, drinking,...*)

I9 SUPER-STRUCTURE MARKER

Inspects a segment for the presence of a superstructure-marker.
If this is present, the function is true; it assigns a value to the features <type> and <token>.
<type=direct>: element refers litterally to a category in the superstructure
<type=derived>: element refers to a category by a reasoning on the basis of world knowledge
<token=introduction>: element is a word or phrase expressing a relation to the introduction (e.g. *I want to resemble*, proper name of protagonist)

<token=description>:	element is a word or phrase expressing a relation to the description-part; it is activated by a reference to the topic in subject position (e.g. pronoun referring to protagonist)
<token=conclusion>:	element is a word or phrase expressing a relation to the conclusion (e.g. *concluding*, proper name of protagonist, ellipsis, meta-statement about the text, reference to writer or assignment, ...)
<token=comment>:	element is a word or phrase expressing an evaluative comment to a preceding statement. The comment may - but not necessarily has to- start with a reference to preceding statement, followed by a reference to writer (*I*), followed by a verb denoting an evaluation, e.g. *which I do not like, I do appreciate that*, ...)

Table 3. Auxiliary variables used in CONNECT

Name	Definition
$Context	Specifies textual context of current segment. Value set: *Characteristic*, *Action*, *Response*. Value is set to *Characteristic* when analysis starts; value changes only when applying certain CONNECT-rules.
$Segment	Specifies segment with antecedent of referential expression in current segment. Value set: *line-number of antecedent-segment*, *zero*. Value is set to *zero* when analysis starts; value changes only when applying INSPECT-rule REFERENT; for each segment value is set to *zero* automatically when processing is completed.
$Time	Specifies that current segment functions as temporal marker for following segment. Value set: *true*, *false*. Value is set to *false* when analysis starts; value changes only when applying certain CONNECT-rules.

Table 4. CONNECT-rules

Name	Condition	Action
Contract-0	Contract=false	Continue with Connector-0
Contract-1	Contract<position=pre+verb>	Subordinate to preceding segment; label as *Specification*
Contract-2	Contract<position=preverb *or* verb> & [Connector=false or Connector<token=additive>]	Coordinate to preceding segment; label as *List*
Connector-0	Connector=false	Continue with Marker-0
Connector-1	Connector<position=medial- *or* final-subclause> & [Connector<token=temporal> or Temporal-marker=true; <token> ne frequency]	Subordinate to preceding segment; label as *Background*; make secondary link to Action/Response-line; label as *Sequence*
Connector-2	Connector<token=alternative>	Coordinate to preceding segment; label as *Alternative*
Connector-3	Connector<position=main clause>; <token=additive>	Collocate to preceding segment; label as *Chain*
Connector-4	Connector<position=initial-subcl>; & [<token=temporal> or Temporal-marker=true]	Subordinate to main clause; set $Time=True; label as *Background*; make secondary link to Action/Response-line; label as *Sequence*
Connector-5	Connector<position=initial-subcl>; <token=purpose>	Open Response-line; set $Context=Response; continue with Referent-0
Connector-6	Connector<position=initial-subcl>; & [Modal-marker=true or Response-marker=true]	Open Response-line; set $Context=Response; continue with Referent-0
Connector-7	In all other cases	Subordinate to main clause; label as <*token*>; continue with Referent-0
Marker-0	Action-marker=false & Modal-marker=false & Temporal-marker=false & Response-marker=false	Continue with Referent-0
Marker-1	Referent<position=preverb>; <token=deictic>	Continue with Referent-0
Marker-2	Temporal-marker<token=past>	Continue with Referent-0
Marker-3	$Context=Action	Continue with Action-1

Marker-4	$Context=Characteristic	Continue with Character-1
Response-1	Response-marker=true	Open Response-line; set $Context=Response; continue with Referent-0
Response-2	Action-marker<token ne action> & [Modal-marker=true or Temporal-marker<token=sequence>]	Coordinate with Response-line; set $Context= Action; label as *Sequence*
Response-3	Action-marker<token=action>	Coordinate with Response-line; label first attachment to Response-line as *Solution* (if Problem) or *Instrument* (if Goal), and following ones as *Sequence*
Character-1	Temporal-marker<token=frequency>	Continue with Referent-0
Character-2	Modal-marker=true & Temporal-marker<token ne sequence>	Continue with Referent-0
Action-1	$Time=true or Temporal-marker<token=sequence *or* punctual>	Coordinate with Action-line; set $Context= Action; $Time=false; label as *Sequence*
Action-2	Action-marker<token=action>	Coordinate with preceding Action; label as *Overlap*; (if Action-line not yet opened, open Action-line and coordinate with it; label as *Specification*)
Referent-0	Referent=false	stop with CONNECT
Referent-1	Referent<token=topic>; <position=preverb>	Coordinate with Characteristic-line; set $Context=Characteristic; label first attachment to Characteristic-line as *Specification* and following ones as *List*
Referent-2	Referent<type=endophor>; <position=preverb>; <ant-pos-in-text=adjacent>	Subordinate with $Segment; label as *Specification*
Referent-3	Referent<type=endophor>; <position=preverb>; <ant-pos-in-text=aligned>	Coordinate with $Segment; if $Context= action, label as *Overlap*, in all other cases label as *List*
Referent-4	Referent<position=verb>; <ant-pos-in-segment=verb>	Collocate with $Segment; label as Chain
Referent-5	Referent<type=endophor>; <position=preverb>; <ant-pos-in-text=remote>	Collocate with $Segment (or to its main clause when $Segment is a subordinated clause); label as Chain

Referent-6	Referent<type=endophor>; <position=post-verb>; <ant-pos-in-text=remote>	Collocate to $Segment; label as *Chain*
Referent-7	Referent<type=word-stem>; <ant-pos-in-text=remote>	Collocate to $Segment label as *Chain*
--	True	stop with CONNECT

Table 5. INTEGRATE-rules

Name	Condition	Action
Superstruc-marker-1	Superstructure-marker <token=introduction>	Open Intro-line; attach segment to it; consider object-phrase as text's topic; label as *Specification*
Superstruc-marker-2	Superstructure-marker <token=description> & Referent<token=topic>	Open Characteristics-line
Superstruc-marker-3	Superstructure-marker <token=comment>	Put label *C* in the connecting-line
Superstruc-marker-4	Superstructure-marker <token=conclusion>	Open Conclusion-line; attach segment to it; label as *Specification*
Theme-marker-1	Theme-marker=true & Segment is hierarchically subordinated in microstructure	Make secondary link to Description-line; put Theme-marker<token> as label on attachment point
Theme-marker-2	Theme-marker=true & Superstructure-marker <token=introduction>	Put Theme-marker<token> as label in the connecting-line
Theme-marker-3	More than one Theme-marker	Put all Theme-marker<token>'s on attachment point with Description-line
Theme-marker-4	Only one Theme-marker	Put Theme-marker<token> on attachment point with Description-line
Check	Linking is discontinuous or segment has an isolated linking to a line in the macrostructure	Take the next outcome of the list which resulted from running through CONNECT (provided it does not have the same or worse consequences)

An example

To illustrate how PISA operates, we give the analysis of the first six segments of the text presented in Chapter 4, table 1 (the PISA structure is presented in figure 3 of the same chapter.) These first lines read as follows:

1 I would
1a if forced to make a choice
1 like to be like Ed Nijpels.
2 He looks quite nice,
3 is dressed well,
4 and expresses himself very well.
5 Besides he has undertaken all kinds of activities alongside his rather varied career.

Before entering INSPECT, the segments are parsed as follows:

	Pre-verb	Verb	Post-verb
1	I	would like to be	like Ed Nijpels
1a	if	forced	to make a choice
2	He	looks	quite nice
3		is	dressed well
4	and	expresses	himself very well
5	Besides he	has undertaken	all kinds of activities alongside his rather varied career

INSPECT returns for each segment the following results:

1 I would like to be like Ed Nijpels
REFERENT <position=pre-verb>; <type=pronoun>; <token=deictic>
MODAL-MARKER <type=auxiliary verb>; <token=desire>
SUPSTRUC-MARKER <type=direct>; <token=introduction>

1a if forced to make a choice
CONNECTOR <position=medial subclause>; <type=subordinator>; <token=condition>

2 He looks quite nice
REFERENT <position=pre-verb>; <type=pronoun>; <token=topic>
THEME-MARKER <type=derived>; <token=appearance>
SUPSTRUC-MARKER <type=direct>; <token=description>

3 is dressed well
CONTRACT <position=pre-verb>
THEME-MARKER <type=derived>; <token=appearance>

4 and expresses himself very well

CONTRACT	<position=pre-verb>
CONNECTOR	<position=main clause>; <type=coordinator>; <token=addition>
THEME-MARKER	<type=derived>; <token=character>

5 Besides he has undertaken all kinds of activities alongside his rather varied career

CONNECTOR	<position=main clause>; <type=adverb>; <token=addition>
REFERENT	<position=pre-verb>; <type=pronoun>; <token=topic>
REFERENT	<position=post-verb>; <type=full NP>; <token=topic>
ACTION-MARKER	<type=main verb>; <token=action>
TEMPORAL-MARKER	<type=verb inflexion>; <token=past>
THEME-MARKER	<type=direct>; <token=career>
THEME-MARKER	<type=derived>; <token=career>

CONNECT and INTEGRATE return for each segment the following results:

1 I would like to be like Ed Nijpels

Marker-1	Referent<position=preverb>; <token=deictic>	Continue with Referent-0
Superstruc-marker-1	Superstructure-marker <token=introduction>	Open Intro-line; attach segment to it; consider object-phrase as text's topic; label as *Specification*

1a if forced to make a choice

Connector-7	In all other cases	Subordinate to main clause; label as *<token>*; continue with Referent-0

2 He looks quite nice

Superstruc-marker-2	Superstructure-marker <token=description> & Referent<token=topic>	Open Characteristics-line
Referent-1	Referent<token=topic>; <position=preverb>	Coordinate with Characteristic-line; set $Context=Characteristic; label first attachment to Characteristic-line as *Specification* and following ones as *List*
Theme-marker-4	Only one Theme-marker	Put Theme-marker<token> on attachment point with Description-line

3 is dressed well

Contract-2	Contract<position=preverb *or* verb> & [Connector=false or Connector<token=additive>]	Coordinate to preceding segment; label as *List*
Theme-marker-4	Only one Theme-marker	Put Theme-marker<token> on attachment point with Description-line

4 and expresses himself very well

Contract-2	Contract<position=preverb *or* verb> & [Connector=false or Connector<token=additive>]	Coordinate to preceding segment; label as *List*
Connector-3	Connector<position=main cl>; <token=additive>	Collocate to preceding segment; label as *Chain*
Theme-marker-4	Only one Theme-marker	Put Theme-marker<token> on attachment point with Description-line

5 Besides he has undertaken all kinds of activities alongside his rather varied career

Connector-3	Connector<position=main cl>; <token=additive>	Collocate to preceding segment; label as *Chain*
Marker-2	Temporal-marker<token=past>	Continue with Referent-0
Referent-1	Referent<token=topic>; <position=preverb>	Coordinate with Characteristic-line; set $Context=Characteristic; label first attachment to Characteristic-line as *Specification* and following ones as *List*
Theme-marker-3	More than one Theme-marker	Put all Theme-marker<token>'s on attachment point with Description-line

Next to showing how PISA operates, the above example also illustrates a number of its characteristics. We round off by mentioning three of them.

Segments with very different INSPECT-results may still be alike in their CONNECT-results. The segments 2 to 5 do not all have a CONNECTOR, CONTRACTION, MODAL-MARKER or TEMPORAL-MARKER, but each of them gets attached to the characteristics-line. Especially the CONNECT-results of segments 2 and 3 make clear that the same result can be obtained via different routes.

A marker may not take effect when co-occurrent with another marker. In segment 1 a deictic reference overruled the MODAL-MARKER, in segment 5 a TEMPORAL-MARKER neutralized the ACTION-MARKER; they made CONNECT go directly to the referent-rules.

CONNECT considers its results in their order of production: the first on the list being the primary candidate. There is only one exception to this decision rule. Additive connectors show up very early in CONNECT, but have only a collocation as result. If later on another marker points at a stronger relation, this attachment becomes first on the list of results. This explains why in segments 4 and 5 the referent-outcome took precedence over the connector-result.

Appendix to chapter 5: the original Dutch text

Table A. Dutch version of the text presented in table 6 of chapter 5

1	Ik zou graag willen lijken op de tenniser André Agassi.
2	Hij is ongeveer even oud als ikzelf
2a	(25 jaar)
3	en verdient zijn brood met mijn hobby,
3a	nl. sport.
4	Hij is dus professioneel sporter
5	en
5a	aangezien hij tot de top behoort,
5	verdient hij een zeer dikke boterham.
6	Hij onderscheidt zich in het tenniswereldje door zijn opvallend en afwijkend uiterlijk en door zijn goede spel natuurlijk
7	Hij komt
7a	ondanks vaak agressieve spel
7	toch zeer sympathiek over
8	en dat mag ik wel in iemand
9	Ik denk niet dat hij veel vrije tijd overhoudt
10	maar
10a	als hij vrij is
10	dan omgeeft hij zich door allerlei luxe
11	en smijt zijn zwaar verdiende geld lekker over de balk.
12	Een lekker figuur
12a	die keihard wil werken
12b	en er toch plezier in houdt
12c	en af en toe de zoete vruchten plukt van zijn geploeter.

Appendix to chapter 6: the original Dutch texts

Table A. The Dutch version of the text presented in table 3 of chapter 6

1	gerad van Tienen.
2	8 jaar
3	hij ziet er goed uit.
4	hij kan middel-madig leren.
5	zijn vader is Timmerman.
6	hij is veel buiten.
7	hij speelt veel.
8	eet wijnig.
9	hij maakt zijn eigen vuil.
10	veel bij andere mensen.
11	klein.

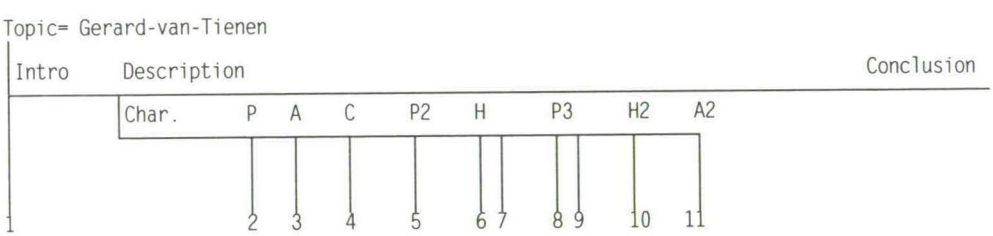


Figure A. PISA structure of the text presented in table A

Table B. The Dutch version of the text presented in table 4 of chapter 6

1	E.P. 's Graafland. Voetballer
2	Van beroep is hij voetballer bij Fijenoord.
3	Keeper
4	Hij kan goed keepen
5	Hij heeft zwart haar
6	en is 1.69 m. lang ongeveer
7a	Als de spelers ruzie hebben
7	komt hij ertussen
8	en is het zo afgelopen
9	Bij de wedstrijd tegen real Madrid in Nederland is het er hard aan toe gegaan.
10	Een Spanjaard werd beentje gelicht
11	en toen was het aan de gang
12	Op de tribunes had iemand met een bierflesje 'n ander op het hoofd geslagen.
13	Hij is vriendelijk
14	en een goede keeper
15	Hij heeft rugnummer 1.

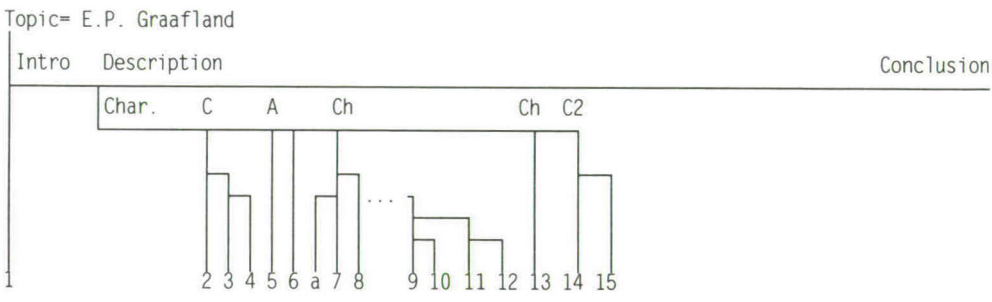


Figure B. PISA structure of the text presented in table B

Table C. The Dutch version of the text presented in table 5 of chapter 6

0	Gymnastiekleraar
1	De persoon waar ik het nu over ga hebben geeft bij ons aan de Ulo les
2	en we zullen hem voor het gemak maar Jan noemen.
3	Jan is een zeer gemakkelijk mens,
3a	die je geen straf geeft,
3b	als je er tenminste niet om "vraagt".
4	Het is een zeer goede leraar
4a	die fijne gymnastiek geeft,
4b	ofschoon het soms ook wel vervelend is.
5	Ik kan er trouwens heel goed mee opschieten.
6	En ik was van plan ook het vak te leren wat Jan geeft.
7	De leeftijd van Jan weet ik niet,
8	maar ik schat hem op ± 30 jaar
8a	wat ik overigens een aardige leeftijd vind voor het geven van dat soort lessen.
9	Het karakter van Jan, is mij niet bekend,
10	maar ik zal er proberen iets over te schrijven.
11	Volgens mij is het karakter rustig,
12	want hij windt z'n eigen niet gauw op
13	ik geloof ook wel dat hij een goed huwelijk heeft, met één kind.
14	Meer zou ik niet kunnen zeggen,
15	ik kan natuurlijk wel gissen,
16	maar daar heeft degene die dit nakijkt, ook niet veel aan
16a	denk ik zo.
17	De figuur van Jan is volgens mij goed
18	de lengte van benen, armen romp en onderlichaam is goed,
19	en het gezicht, met de blauwe ogen en het zwarte haar is ook niet slecht.
20	Het beroep van Jan is gymnastiekleraar
20a	zoals enkele regels in het begin al vermeldden
21	Zijn ontspanning zoekt hij volgens mij in de tuin en in de huiskamer (of buiten)
21a	waar hij leest in een of ander boek
21	en op het rijden van zijn scooter
21b	(wat ik overigens flauwe kul vind)
22	Dit is dan mijn relaas over de gymnastiekleraar van onze school

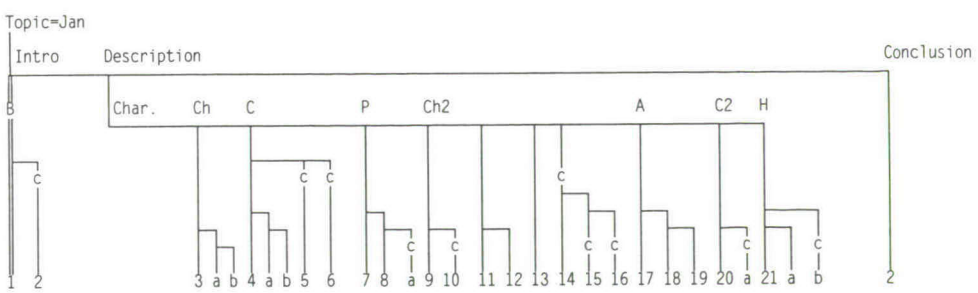


Figure C. PISA structure of the text presented in table C

Curriculum vitae

Els van der Pool behaalde in 1983 haar gymnasiumdiploma aan het St.Oelbertgymnasium in Oosterhout. In september van dat jaar begon zij haar studie Taal- en Literatuurwetenschap aan de Katholieke Universiteit Brabant in Tilburg. In 1989 behaalde zij het doctoraal examen met als specialisatie Tekstwetenschap.

Tijdens en na haar studie werkte zij als projectmedewerker aan een kantoorautomatiserings-project bij de Sociale Dienst in Tilburg.

Vanaf 1990 was zij werkzaam als Assistent-in-Opleiding bij het werkverband Tekstwetenschap van de Letterenfaculteit aan de Katholieke Universiteit Brabant. Daarnaast werkte zij als freelance trainer en onderzoeker voor diverse opdrachtgevers, waaronder de Belastingdienst/Centrum voor personeelsontwikkeling en organisatie-advies.

Vanaf 1 oktober 1995 is zij werkzaam als trainer/adviseur schriftelijke communicatie bij Boertien & Partners, bureau voor advies en training, in Naarden.

Stellingen behorende bij het proefschrift

Writing as a conceptual process.
A text-analytical study on developmental aspects

* * * * *

1 In de conceptuele component van het schrijfproces moeten drie operaties worden onderscheiden, aan te duiden als *Reflecteren*, *Selecteren* en *Vormgeven*. (hoofdstuk 2)

2 Syntactische kenmerken hebben een relatie met zowel talige als conceptuele operaties. Daardoor geven ze minder duidelijk zicht op conceptuele processen dan structuurkenmerken. (hoofdstuk 3)

3 PISA is een methode om tekststructuren te analyseren op basis van geëxpliciteerde regels en kennis. Het instrument maakt directe vergelijkingen tussen teksten mogelijk en levert bruikbare data op voor psycholinguïstisch onderzoek. (hoofdstuk 4)

4 De kwalificatie "in der Beschränkung zeigt sich erst der Meister" geeft precies weer dat niet alle processen van vaardige schrijvers zich manifesteren in één tekstsoort. (hoofdstuk 5)

5 Op conceptueel niveau wordt de ontwikkeling tot een vaardig schrijver gekenmerkt door een toename van *Reflecteer*-operaties. Dit kan globaal worden omschreven als de overgang van procedureel naar strategisch werken. (hoofdstuk 6)

6 Voor een theoretische verdieping van schrijfprocesonderzoek is het noodzakelijk het onderzoeksdomein niet te beperken tot het woord- en zins-niveau maar uit te breiden met het tekstniveau. (hoofdstuk 7)

7 Schrijftrainingen dienen gericht te zijn op het aanleren van strategieën en niet beperkt te blijven tot het oefenen van procedures.

8 Wil schijn niet bedriegen dan zou deze alleen mogen worden opgehouden indien alle betrokkenen er bewust naar handelen.

9 Evenwichtige verdeling van (onbetaalde) zorgtaken is pas te realiseren als zowel 'moeder-de-vrouw' als 'vader-de-man' respectabele en na te streven kwalificaties zijn.

10 Een assistent-in-opleiding zal na de opleiding zelden assistent zijn op de universiteit.